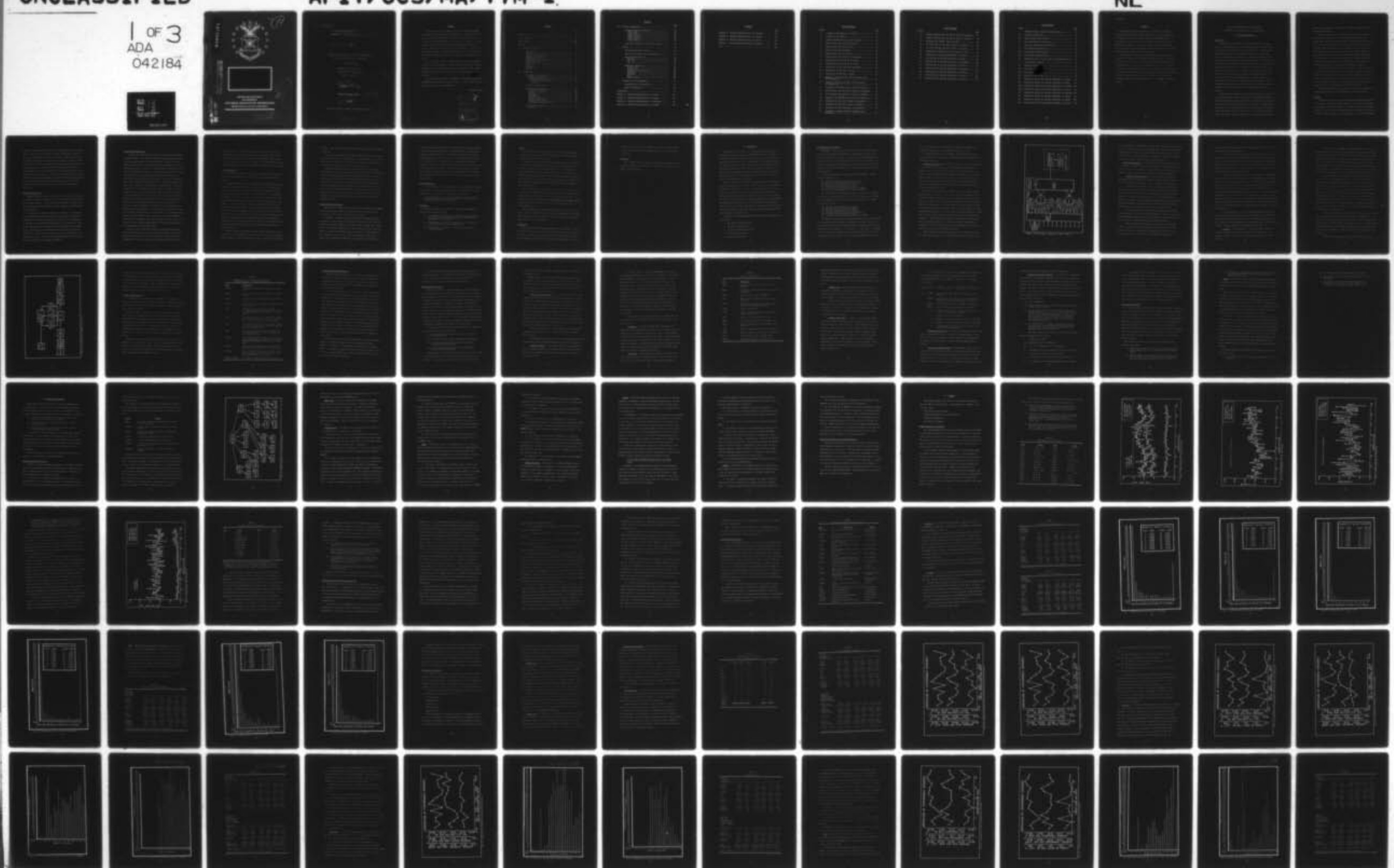
AD-A042 184 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO SCH--ETC F/G 9/2
WORKLOAD CHARACTERIZATION AND MEASUREMENT OF THE CDC CYBER 74 C--ETC(U)
MAR 77 J R BEAR

UNCLASSIFIED AFIT/GCS/MA/77M-1 NL

1 OF 3
ADA
042184



UNITED STATES AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

(14) AFIT/GCS/MA/77M-1

(6) WORKLOAD CHARACTERIZATION AND

MEASUREMENT OF THE CDC CYBER 74 COMPUTER SYSTEM

(9) Masters THESIS,

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

by

(10) Jonathan R. Bear, M. S.
Captain USAF

Graduate Computer Systems

(12) 234 p.

(11) March 1977

Approved for public release, distribution unlimited

Preface

This thesis topic comes within the area of computer performance and evaluation, and involves an investigation of the workload upon the Control Data Corporation CYBER 74 computer system. Several days of computer system accounting data were processed. The workload was characterized by use of a variety of statistical measurements, and the impact of variations in workload upon the performance of the computer system was measured. The methodology and results of this thesis could be used as a preliminary step in an evaluation of the CYBER computer system.

I wish to thank a few of the many individuals who have generously helped in this investigation. I wish to thank the ASD Computer Center personnel for their continued support. For providing me with valuable assistance and direction, I also wish to thank Dr. Ed Reeves, my advisor.

My primary thanks must go to my family for their encouragement and understanding throughout the course of study. I would especially thank my wife, Cheryl, for her contributions.

Jonathan R. Bear

Contents

Contents

Contents

List of Figures

List of Figures

<u>List of Tables</u>

Abstract

In evaluating computer performance, the choice of a specific evaluation approach depends upon the objectives of the investigation, although basically, the methodology involves the observation, and (observing! measuring the measurement of a computer system or its model while a set of jobs, or workload, is processed. Computer performance evaluation, then, can be divided into 2 two focal points; the computer system and its workload. The focus of this thesis ~~is~~ focuses upon the workload.

A methodology was developed to process accounting data generated by the Control Data Corporation CYBER 74 computer system. A number of statistical tools were used to measure several representative days of accounting data. The workload was characterized by use of a variety of statistical techniques, and the effect of variations in the workload upon computer system performance was measured. The methodology developed for this investigation and the results obtained could be used as a preliminary step in a performance evaluation of the CYBER 74 computer system.

WORKLOAD CHARACTERIZATION AND PERFORMANCE MEASUREMENT OF THE CDC CYBER 74 COMPUTER SYSTEM

I. Problem Specification

Introduction

Measurement and evaluation of computer performance is an area which is growing rapidly due to the economic necessity to derive maximum utilization from computer facilities (Ref 25:3). Applications of performance evaluation in the computer science field are many and diverse. For example, the evaluation of computer performance could be involved with the selection of new computer systems, with the design of applications and equipment, and with the examination and improvement of existing systems (Ref 20:79). The choice of a specific evaluation approach will depend upon the objectives of a particular study, but, basically, the methodology for a computer performance evaluation involves the observation and measurement of a computer system or its model while a set of jobs, or workload, is being processed. Computer performance evaluation, then, can be divided into two focal points, the computer system and its workload. The focus of this thesis is the workload.

There are various definitions of workload; however, to remain consistent with the objectives of this study, the term workload refers to the specific computer programs which must be processed by the computer system in order to satisfy user requests (Ref 13:3). Workload may be contrasted to other software programs, such as compilers and operating system algorithms, which are considered to be part of the system overhead. The computer system is defined as the physical machine itself,

all associated peripheral equipment, the workload, overhead, and the man-machine interface.

Due to the expense and time involved in measuring the workload of an existing computer system, several authorities, such as Watson, Sceenivasan, and Esposito, have suggested a cost-effective approach to this problem (Refs 29:1; 27:6; 12:3). Such an approach would involve computer system accounting data, the information which is collected about resource usage in order to charge cost to users. Analysis of such data could be a valuable tool in assessing computer performance, the effectiveness with which the resources of a computer system are utilized. Since the data is already available, this approach also offers the advantage of being relatively inexpensive (Ref 29:1).

In this thesis, accounting data is processed in order to characterize the computer system user's workload and to measure the impact of variations in workload upon the computer system performance. The results of this analysis could be used as an initial step toward a complete performance evaluation. The approach and terminology will be further developed in subsequent sections of this thesis.

In order to familiarize the reader with the area of computer performance and evaluation, a brief history will first be outlined.

History

Kimbleton (Ref 19:2), in his text discusses an abbreviated history of computers. The rather late development of the computer performance and evaluation area may be related to the rapid development of computer systems. Since the first and second generation of computer systems processed only one program at a time, the concern was with the performance

measure of jobs processed per unit of time (throughput) and the processing time for each job (turnaround). System performance was related to how quickly an individual program was processed. When the early multi-programmed third-generation computers were developed, the main concerns were reliability of the operating system programs and reliability of the physical computer. However, with further technological advances in third-generation computers, evaluation of computer performance on a more complex level became essential. Since multi-programmed computer systems processed users' jobs simultaneously by distributing the system resources to various program requests, an evaluation approach which took into account the workload and its effect on the system became necessary (Ref 19:13).

Performance Definition

Although there has been much study both on a theoretical level and on a practical level, a comprehensive theory of computer performance has not yet been developed. Little progress has been made in solving very basic problems, such as acceptably defining performance and characterizing the workload.

In this thesis, computer performance is defined as the effectiveness and efficiency with which a computer system uses its resources to accomplish the computer system objectives (Ref 28:8). Performance may be measured by such quantitative parameters as (1) utilization of individual components, such as central processor utilization; (2) internal delays, such as the time between submitting a job into the system and receiving the output; and (3) productivity, such as the number of jobs processed per hour. A variety of performance measures may be used with this broad definition of computer performance.

Workload Characterization

A variety of techniques may be used to evaluate the performance of a computer system. Such techniques include analytical modeling, simulation, and experiments on the existing system. Since all of these techniques focus upon the interaction between the workload and the computer system, there is a need for a representative workload that simulates with reasonable accuracy the actual workload (Ref 27:127). Since the actual workload of a multiprogrammed computer generally is not reproducible and is extremely variable, there are several advantages to using a representative workload (Ref 3:193). Among the advantages of using such a workload are its stability, reproducibility, flexibility, and brevity. The representative workload should be stable and reproducible in order to maintain consistency during performance evaluation; it should be flexible in order that the characteristics of the representative workload may be varied (Ref 26:4). This is necessary since there can be a wide variation in actual workloads (Ref 3:193). In addition, it may be necessary to represent the wide variations in the actual workload by separating it into different classes based upon specific workload characteristics, as discussed below. A potential pitfall, however, in simulating the actual workload is that the representative workload may not adequately represent the actual workload; it may not be possible to design certain characteristics of the actual workload into the representative workload (Ref 30:53).

Since the actual workload can serve as the basis from which the representative workload is derived, the actual workload must be characterized quantitatively. In this thesis, several workload classes are characterized by the type and amount of resources which are used by a

computer system to satisfy requests from users' programs (Ref 28:10). For example, some workload characteristics are central processor unit (CPU) time, input/output (I/O) and memory requirements. The values of these job parameters can be used to quantify the representative workload by analyzing the distribution of demands upon individual system resources.

Measurement Tools

A variety of tools have been used and are available to measure computer system performance; the most common are hardware and software monitors.

Hardware monitors consist of a collection of high-impedance probes, which are connected to the computer circuitry, and a recording device. More complex monitors may also have a computer to preprocess and/or analyze the data which is gathered. Upon activation, a probe generates a signal which is sent to a timer or counter. The results may be used to determine how the hardware responds to computer system requests. An example of such use would be a hardware monitor measuring the time in use of an input/output channel. Hardware monitors offer two important advantages: they obtain precise data without affecting the computer system, and they may be activated and deactivated without affecting processing rates. Disadvantages of hardware monitors include the relatively small number of probes normally used, the difficulty of tracking the computer resources used by individual jobs, and the relatively high cost involved (Ref 5:14).

Software monitors are computer programs which usually reside in the central memory of the computer; they are used to collect data on computer system operations, such as queue activity, CPU utilization, and supervisor

routines. Such data may be used later as part of a computer performance evaluation.

The primary disadvantage of software monitors is that they consume a portion of the computer resources, and thus, detract from the system's performance (Ref 23:31). As a result, unless carefully designed, software monitors can distort the data they are measuring.

Accounting systems, which are used on virtually all large scale third-generation Air Force computers, are among the more common software monitors. Installed primarily to charge users in proportion to their use of computer resources, the accounting systems usually contain historical data in the form of event-oriented parameters. For example, these data generally include information about system resources used by each job rather than information about system resources used by the computer system as a whole. As a result, these data are more suitable for evaluating workload characteristics than for evaluating computer performance (Ref 29:4).

Investigation Justification

Increased user demands upon the Control Data Corporation (CDC) CYBER 74 have resulted in a number of complaints by the users of the system. The history of this problem will be briefly discussed.

The CYBER 74 computer system was installed at Wright-Patterson Air Force Base in 1971; problems and complaints from users of the system were initially handled on a weekly basis at Happy Users Meetings (HUM). After most of the operating system flaws had been corrected, the HUM were held monthly. During the intervening years, the number of programs processed daily increased from 300 to 1700. An interview with Mr. Jim Hudson, Chief of the Program Control Branch at the Aeronautical Systems Division

(ASD) Computer Center, indicated that during 1976 the number of service complaints had increased markedly. These complaints, involving slower turnaround time for batch jobs and decreased response time on interactive terminals, resulted in a series of special meetings which were held in April of 1976. Participants in these meetings suggested a variety of solutions to return the performance of the system to the level formerly enjoyed by the users. However, no viable solution was found, and the final conclusion of the meetings was that the actual workload was unknown and the effect of the workload on the computer system was also unknown.

Problem Statement

In this thesis, it is proposed that a study of the CYBER 74 computer system could be a step toward answering the questions described above; specifically, the problem to be addressed is as follows:

> To characterize the workload, and to measure the effects of variations in workload upon the performance of the Control Data Corporation CYBER 74 computer system.

Objectives

Several objectives were formulated to guide the investigation. These objectives are as follows:

(1) To investigate the adequacy of data from accounting systems as a basis for characterizing workload and measuring the performance from that workload.

(2) To determine the workload characteristics for a variety of actual workloads from the accounting data.

(3) To investigate the effects of variations in workload on system performance.

Scope

This study focused specifically upon the system accounting data collected from the Control Data Corporation CYBER 74 computer system located at the ASD Computer Center.

A large amount of accounting data is gathered by a CDC accounting system program, Dayfile, which computes a variety of parameters for each job. The accounting data are generally used to calculate cost to users, but the Dayfile also contains such parameters as memory used, execution time, and input/output time. The available Dayfile accounting data are stored for a six month period and are available on magnetic tape. These data are a valuable source of information for an investigation of the computer system.

Due to restrictions in available computer time for this thesis, this investigation was further limited to magnetic tapes containing data from the software program Computer Load and Resource Analysis (CLARA). Designed by the Boeing Computer Service, Incorporated, this program reduces Dayfile data, often containing as many as 60,000 records per day, to a more manageable size.

Since accounting data were the major source of information for this thesis, the focal point of observation is the job and system level rather than the machine or circuit level. For the same reason, the man-machine interaction will not be investigated.

Constraints

Data gathering for this investigation had to be restricted to information sources which were readily available and which met cost restrictions for thesis projects. Due to the limited time for development, installation, and analysis, new hardware monitors and software monitors were not

purchased. There were no other available hardware or other software monitors which were readily accessible within the time constraints of this thesis.

Assumption

It is assumed that the workload characterization generated from the accounting data will be representative of the actual workload on the CYBER 74 computer system.

II. Methodology

The purpose of this chapter is to discuss the methodology used to accomplish the thesis objectives. The methodology follows several of the Rand Corporation's suggested initial procedures for an evaluation of computer performance (Ref 4:10). Some departures from these procedures were necessary since this thesis focuses upon workload rather than upon performance measurement. Suggestions and techniques offered by such authors as Watson, Agrawala, Dodson, and Gudes were also used as guides (Refs 29, 1, 10, 15). The following steps were taken on an iterative basis while developing the methodology.

First, an analysis of the target computer system and normal operations was necessary before one could proceed with further investigation. An analysis was required to select the workload characteristics; then, statistical techniques were selected to quantify these measurements. Several performance measurements were also selected in order to measure the effects of various workloads. Finally, an analysis of the available accounting data was completed in order to ensure that the appropriate parameters were present for investigation. Since hardware and other software monitors were not available, workload and performance measures were modified based on available data.

A summary of the various analyses developed in this methodology are listed below:

1. CDC CYBER 74 Computer System
2. Workload Characteristics
3. Workload Measurement Techniques
4. Performance Measurements
5. Data Collection System

CDC CYBER 74 Computer System

The CYBER 74 computer is structured as a part of a larger computer network, the Aeronautical Systems Division (ASD) computer system, which also contains a CDC 6600 computer. At selected times, the CYBER 74 can process a portion of the workload from the CDC 6600, and vice versa. In addition to the central site facilities, the ASD computer system has the capability for interactive access by 40 teletype terminals and 9 remote batch terminals.

The organizations which can directly access the CYBER 74 computer system are listed below:

(1) .Air Force Institute of Technology (AFIT)
(2) Air Force Flight Dynamics Laboratory (AFFDL)
(3) Air Force Avionics Laboratory (AFAL)
(4) Aeromedical Research Laboratory (AMRL)
(5) Air Force Wright Aeronautical Laboratory (AFWAL)
(6) Air Force Human Reliability Laboratory (AFHRL)

Jobs of several other organizations may be processed by the CYBER 74 if the jobs are transmitted from the CDC 6600 for execution on the CYBER 74. The organizations which can directly access the CDC 6600 computer system are as follows:

(1) Aerospace Systems Division/AD (ASD/AD)
(2) Aerospace Systems Division/EN (ASD/EN)
(3) Aerospace Systems Division/XR (ASD/XR)
(4) Aerospace Systems Division/other (ASD/other)
(5) Air Force Materials Laboratory (AFML)
(6) Air Force Aerospace Propulsion Laboratory (AFAPL)

The CYBER 74 normally operates continuously with all its remote batch and teletype terminals; however, an occasional backlog of jobs at the batch terminals has forced the Computer Center to disconnect the teletype terminals for short periods. In an attempt to provide adequate turnaround time for smaller jobs, the policy of the Computer Center is to limit the central memory field length to 120,000 octal words between the hours of

0800 to 1630 (Ref 2:9). In addition, the interactive terminals and remote batch terminals are not operative from 2400 to 0800.

A general description of the CYBER 74 computer system follows: the discussion is divided into two categories, hardware architecture and software architecture.

Hardware Architecture. The CYBER 74 computer, a more recent variety of the earlier CDC 6000 series, is a multiprogrammed file-oriented machine which can be divided into three parts as seen in Figure 1: a single central processor (CPU), 20 peripheral processors, and a central memory of 131K 60 bit words. The CPU has a high speed arithmetic and control unit which fetches, decodes and executes instructions sequentially. The average execution time for each instruction (major cycle) is one microsecond (Ref 10:13).

Each peripheral processor unit (PPU) is a smaller general purpose computer having 4K memory of 12 bit words. The PPU's purpose is to complete slower system tasks such as I/O and supervisory operations, thus allowing the CPU to carry out high speed computations (Refs 14:33; 6:23). Each PPU can communicate with central memory, and since PPUs do not execute complicated problems, they can operate in a time sharing mode without causing system degradation. The PPUs share common data channels to communicate with the central memory and peripheral devices in a multiplexing arrangement. The arrangement consists of a scheduler, or barrel, with a position representing each PPU. The barrel schedules each position sequentially to receive a time slice of 100 nanoseconds (minor cycle) to execute a portion of an instruction from a PPU.

Central memory and PPU memory consist of 3D core memory sectioned into basic memory modules of 4096 twelve bit words with an access cycle

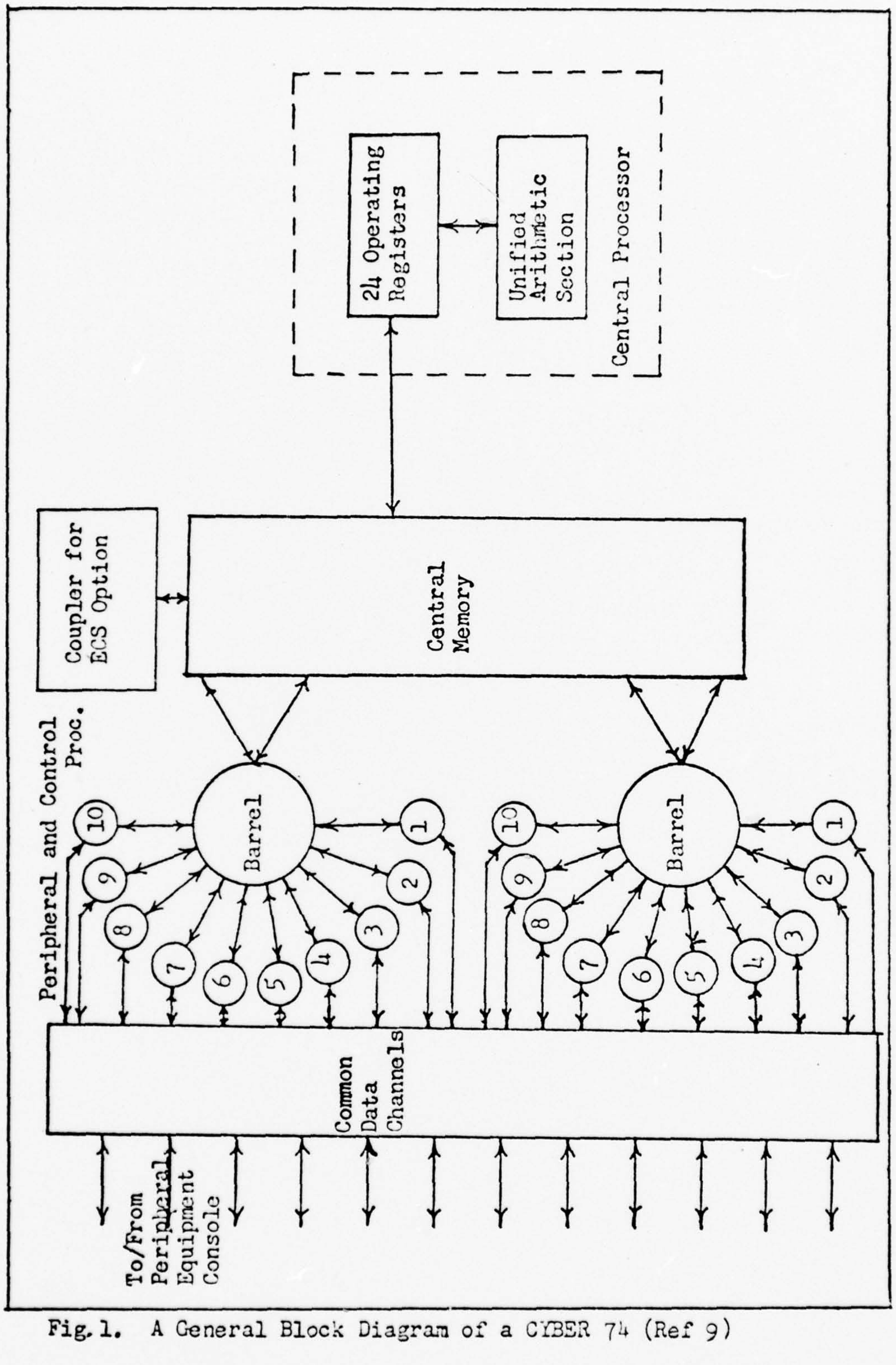


Fig. 1. A General Block Diagram of a CYBER 74 (Ref 9)

of one microsecond. Each PPU has a single module of 12 bit words; however, for the 60 bit words of the central memory, five modules are horizontally connected into a bank, with 32 banks forming the central memory of 131K words (Ref 6:24). Provisions are also available to increase core memory capacity by adding an Extended Core Storage (ECS) option.

Software Architecture. In this section, the software architecture of the CYBER 74 computer system will be outlined. A description of the operating system and its functions will be followed by discussion of the Scheduler, which is an important component of the operating system.

CYBER 74 Operating System. The New Operating System/Batch Environment (NOS/BE) monitors the CYBER 74 computer system. A variety of modes of service are supported: batch, interactive, and graphics.

A system monitor, JANUS, which is located in one of the PPUs, schedules the total operation of the system, including CPU and PPU requests. Since there are no hardware interrupts in the system, this PPU schedules all requests for system resources. The remaining PPUs are assigned a variety of tasks in executing system and user jobs (Ref 7:115). For example, INTERCOM, which controls the scheduling of all interactive terminal requests, is another system program located in a PPU.

In order to provide a multiprogramming capability, the CYBER 74 can simultaneously store 15 jobs in central memory. These jobs reside in 15 variable-length partitions, or control points; the size of a control point is called its central memory field length (Ref 8:2-10).

Although a maximum of 15 active control points are available at any one time, fewer control points are actually available for user jobs since a number of system support programs occupy several control points. JANUS occupies one control point, INTERCOM occupies two,

graphics one, and two control points are occasionally used for communication with other computer systems.

The 15 jobs holding control points can compete for the use of the CPU, although only one job can actually be executed by the CPU at any one time. A control point may be in one of five states: (1) executing with the CPU, (2) waiting for the CPU, (3) waiting for a PPU activity to complete, (4) waiting for some operator action, or (5) rolled out (discussed in Scheduler section) (Ref 7:115).

The priority given a job depends upon what stage of processing it is in. Initially, a job receives a priority depending upon its type. Batch is the lowest in the hierarchy with a priority of 1000. All interactive job commands and requests for execution have a 3000 priority. Graphics has the highest priority which is 6000; an exception is that operator assigned priorities may range as high as 7000. Once a job enters the CYBER 74, it passes through several stages before execution is completed; at each state, priorities are increased by an "aging factor" which is based upon the amount of time spent waiting. In this way, lower priority jobs can eventually compete on an equal basis for execution with jobs of the same type.

It is interesting to note that although INTERCOM and graphics hold only three of the ten available control points for jobs, they could use significantly more than 30 per cent of the CPU's time, since they have a higher priority than batch. As a result, interactive and graphics requests can and do affect the performance of batch jobs.

Scheduler. The Scheduler is a support software program which plays an important role in processing the flow of users' jobs. The Scheduler controls the priority of jobs and the selection of jobs for execution.

The job flow of the CYBER 74 NOS/BE operating system is diagrammed in Figure 2. All batch jobs enter the system from card readers, disks, magnetic tapes, or from the batch mode of INTERCOM. After being assigned a priority, the jobs are stored in the input queue as files on disk. All INTERCOM jobs and graphics jobs are stored in the central memory queue. When a control point is available, the Scheduler selects from the input queue or central memory queue, the program with the highest priority which can fit into the available central memory space.

When a suitable job is found, the Scheduler places the new job into a control point, "swaps in" the job, by first checking to see whether there is sufficient field length for the job. If not, the job with the lowest priority is "swapped out" or loses its control point to make room for the new job. In order that a job can continue execution from the point where it was interrupted, each control point maintains 200 octal words of memory. These words contain such information as register contents and a program counter. When a job is swapped back in, this information is used to continue execution. The Scheduler can monitor up to 99 jobs at any one time, 15 of which can be actively competing for execution time (Ref 22:750).

When a job requests I/O from a magnetic tape, the job is "rolled out" into the central memory queue rather than being swapped out. A "rolled out" job does not lose its control point. When the operation is completed, the job again competes for execution time. As a result, jobs in the central memory queue consist of rolled-out jobs and swapped-out jobs, plus those INTERCOM and graphics jobs which have arrived for the first time.

Only one job of those active jobs having control points can

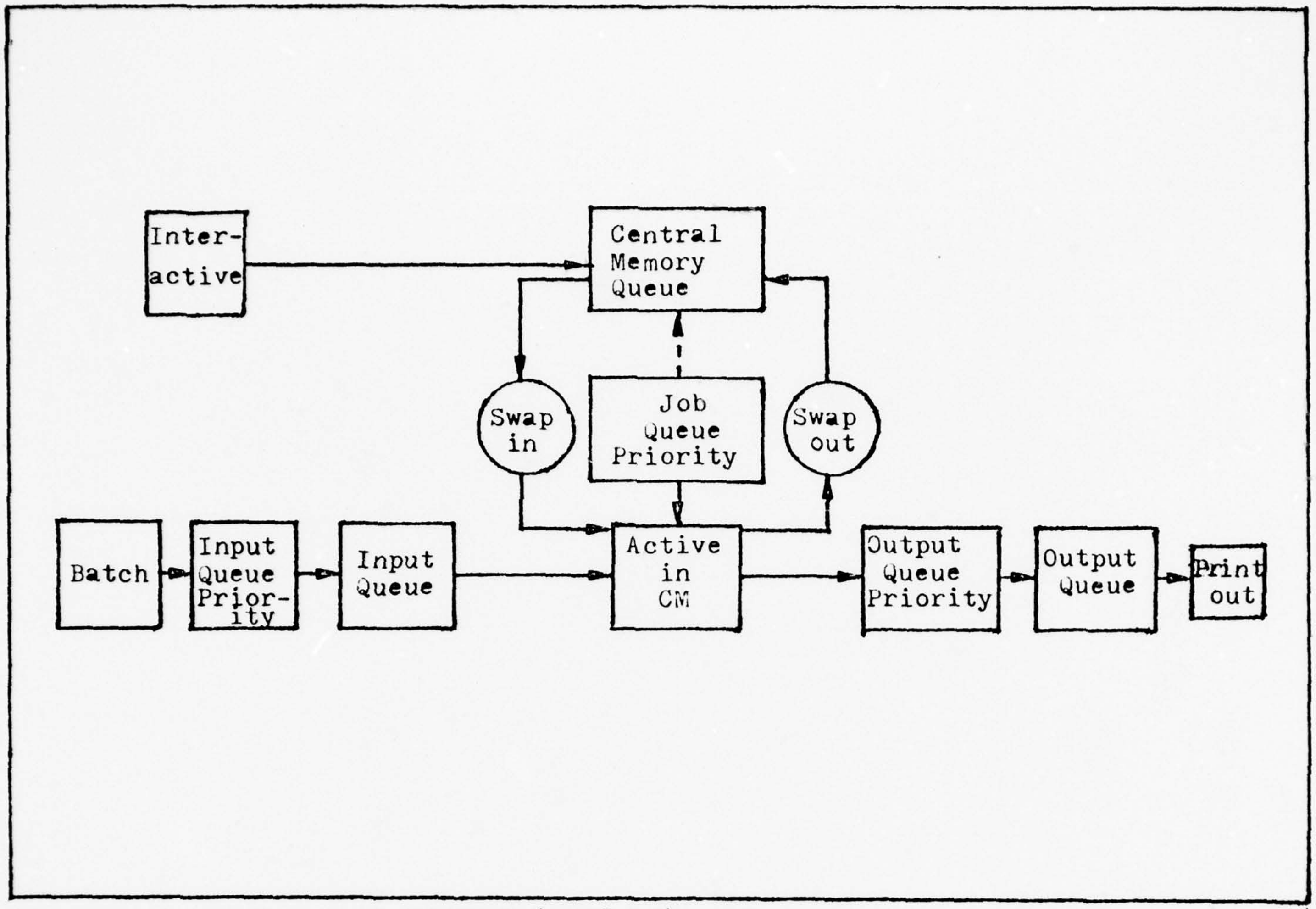


Fig. 2. Job Flow of the CYBER 74 (Ref 8:1-2)

actually execute while in the central memory. The job with the highest priority is given a time quantum in which to execute; at the end of that time, a search for the job with the next highest priority is initiated for the following time quantum. When a job has received sufficient time quantums to complete execution, its file enters the output queue, is assigned a priority, and waits for output to the user.

Workload Characteristics

The methodology for characterizing the CYBER 74 workload uses several statistical techniques, discussed later, to measure and categorize the accounting data by type and amount of resources. This investigation uses a two step approach.

First, each day of accounting data is characterized by resource usage; a number of variables are selected from the accounting data as representative of the actual workload. These variables, reflecting the suggestions of several authors, can be appropriate for a workload characterization (Refs 12:5; 27:7). The parameters in Table I were used to characterize the workload. Even though many other variables were also available, the reasons for their exclusion are presented in the Results chapter.

Second, the variations in the workload among the different days are categorized by separating the days of data into three classes based upon CPU time, I/O time, and memory usage. Although other classes could be formed, this investigation focuses upon these three classes as an initial effort to characterize the CYBER 74 workload.

TABLE I

Parameters Used in Workload Characterization
(All measurements of time are in seconds)

Name	Description
CPTIME	The central processor time needed to process a job.
PPTIME	The peripheral processor time needed to process a job.
TIMEIO	The total input/output time needed to process a job.
TOTCOST	The total job cost in terms of Computer Resource Units (CRUs).
KWS	The memory usage in terms of kilo-word seconds. This variable must be calculated from the central memory term (CMTERM) in the CLARA DATA.
CPOT	The total control point occupancy time which is the time a job enters a control point until it leaves a control point for the last time.
CMLOC	The number of central memory locations occupied by all jobs including this job at the time this job left a control point.
CPLOC	The number of control points occupied by all jobs including this job at the time this job left a control point.
ROLLOUT	The total time that a batch job was rolled out for processing magnetic tapes. This variable is derived from LSTROLL and TOTROLL in the CLARA DATA.
IATIME	The total time between this batch job and the last batch job to arrive into the Input Queue.
INQTIME	The time a batch job spends in the Input Queue. This is derived from the entry time into this Queue (INPUTQ) and its exit (INCP).
TAPEREQ/TAPEUSED	The number of tape requested/used per job.

Workload Measurement Techniques

Several statistical techniques were chosen to characterize the workload of the CYBER 74 computer system. In order to ensure that workload parameters would be adequately analyzed, the statistical techniques which were selected were recommended by several authors (Refs 24:46, 21:16, 12:3).

Simple statistics measure a single aspect of distribution. The mean or weighted average indicates the point of central tendency of a distribution. The mean should be interpreted carefully since a few very high or very low values will shift the mean in the direction of those values out of proportion to their representation in the population. Variability is measured by the range, deviation, and variance. The range is the difference between the highest and lowest observations. The deviation is found by adding together the squared differences between the mean and each observation, dividing by the number of observations, and taking the square root. Although it has the same limitations as the mean, the deviation gives a good indication of the variability of a population. The variance is computed in the same manner as the average absolute deviation except that the square root is not taken.

Graphical plots are used to consolidate data in an easily readable fashion. Frequency histograms are used to illustrate the usage pattern of a variable or its frequency of occurrence over a range of values. In addition, a time-series pattern is plotted in order to display variable usage through time. The time series plots are used as an indication of workload direction.

The joint probability density function is approximated for the workload parameters. The results of this statistical technique could be used to construct a representative workload as described by Screenivasan for use with a computer model (Ref 27:11).

Performance Measurements

Measuring the effect of workload variations upon the performance of a computer system is a complex task because there is no generally accepted measure of performance. The main difficulty can be traced to the absence of acceptable criteria upon which the performance can be optimized (Ref 19:19). One approach to selecting an appropriate measure(s) of performance involves investigating the mission or operational objectives of an organization (Ref 4:10). An effort should be made to select a performance measure consistent with those objectives. However, a problem exists when goals of the organization are broad and not specifically stated: finding a performance measurement(s) which is universally acceptable may be difficult or impossible.

The overall mission of the Aerospace Systems Division (ASD) computer center is stated below (Ref 2:2).

(1) To provide centralized computer support to the AFSC/AFIT organizations.

(2) To develop improved computer techniques and procedures.

(3) To program and operate the general purpose digital, analog, and hybrid computers.

(4) To act as the ASD focal point for computer support, acquisition, and utilization.

The ASD mission as stated above is relatively broad, and it would be exceedingly difficult to select a single adequate measure of performance. It was decided to use a variety of performance measures

since these would better represent the system and be consistent with organization goals.

Several authors have devoted attention to developing measures which might be useful in investigating computer performance. These measures, which are discussed below, were selected in order to measure the effects of various workloads upon the computer system (Refs 29:37; 12:31; 18:23, 11:254; 28:16).

CPU and PPU Utilization. The utilization of the CPU and PPU indicates the percentage of time during which each resource is doing useful work. CPU utilization is defined as the time ratio in which the CPU is active compared to the time that it is available. This ratio is computed by adding the total CPU time for all jobs during the period of investigation, and then dividing this figure by the total time of the investigation period. The PPU utilization is computed in a similar manner. Unlike the CPU utilization ratio, the PPU utilization ratio may occasionally exceed 100% since there will often be more than one PPU active in support of a user's job.

The CPU utilization ratio can indicate when a system is CPU bound; that is, when the CPU is so busy as to create a bottleneck in the job flow.

Since the PPU is generally used for I/O operation, the PPU utilization ratio can be an effective measure of use of I/O facilities.

Overhead Utility. One important method of investigating the workload involves computing the difference between jobs submitted by users, and the overhead, or jobs run in support of and in control of the computer system.

In order to arrive at a better understanding of overhead, two working definitions of this term will be discussed. True overhead has been defined as the software jobs or code necessary to process and control users' jobs in the computer system. Included in this definition would be the compilers, operating system (NOS/BE), the scheduler, swapping algorithm, and priority algorithms. Composite overhead, a working definition used by the Computer Center, has been defined as the true overhead plus those jobs which are executed to help with the maintenance, recordkeeping, and updating of the computer system. Composite overhead, as described in an interview in December of 1976 with Mr. Lou Venuti, a member of the ASD Program Control Branch, includes the Dayfile system, engineers' file, and port reports.

For purposes of this thesis, those jobs in Table 2 are included in the composite overhead investigation on the CYBER 74 computer system.

Throughput. As defined by Hellerman, throughput is the reciprocal of the average time to complete all job requests in a job stream; thus, the throughput indicates the number of jobs completed per unit of time for a given workload (Ref 17:103). Unlike the figure of merit, throughput is insensitive to the order in which the job requests are executed. A system that processes long job requests before short job requests may have the same throughput value as the system which gives priority to shorter job requests.

Turnaround. Turnaround has been defined as the time from when a batch job enters the computer system to the time when output from the job is received (Ref 25:16). This measure includes the

TABLE II

Composite Overhead Jobs

Name	Description
JANUS	A program for control of all I/O in the CYBER 74.
DSTRT	A program for initiating computer operation from a deadstart.
EIEII	Editor job for control of INTERCOM.
B1100	Dayfile dump occurring approximately every two hours.
PORTO	Port reports produced at each remote batch terminal approximately every hour.
DAYFD	Dayfile information.
XMTPT	Job for routing IBM print jobs between the CYBER 74 and IBM 370.
DCDOO	Dump file for the engineers' file.
CEFOO	Additional information for engineers' file.
MFSTA	Control communication between CYBER 74 and CDC 6600.
ASTAT	Control communication link with the CDC 6600.
BSTAT	Control communication link with the CYBER 74.
DMPQO	Job for dumping jobs to magnetic tape.
TIELINE	Communication control with the IBM 370.

man-machine interaction, which is outside the scope of this investigation. Therefore, in this thesis, turnaround time will represent the machine turnaround time, which is the time between a batch job entry into the input queue and the last time the job leaves a control point.

Response Time. An important measure of performance for interactive terminals is response time. This measure is the time elapsed between the sending of the last input character of an interactive command to the time when the first output letter is received.

An investigation of this measure would be important in determining how long a user must wait, and in determining the possible solutions to relatively slow response time. However, due to the lack of appropriate data from the CYBER 74 accounting system, response time could not be investigated.

Economic Performance. A measure of economic performance, and general computer performance, can be derived from the Computer Resource Utilization (CRU) algorithm. The ASD Computer Center uses a measure of resource usage, the CRU, to charge users for their resource usage. The CRU gives a general measure of resource usage since it is a weighted measure of CPU time, I/O time, and central memory occupancy time. The total cost of a job is calculated by charging $0.06 for each CPU generated by a user's program. A more general description of economic and accounting measures is discussed in Reference 16.

The Computer Center algorithm for computing CRUs is described in an ASD letter dated 14 June 1976, "Charges for Use of the ASD Computer Center's CDC 6600 and CYBER 74 Computers." The algorithm is as follows:

$$\text{Number of CRUs} = a_1 \text{CPTIME} + a_2 \text{TIMEIO} + a_3 \sum_{i=1}^{n} (\text{CPTIME}_i + \text{TIMEIO}_i)\text{CM}_i$$

where CPTIME = the total CPU time needed to process the job in seconds.

TIMEIO = the total I/O time used by the job to include total tape channel and disk access time in seconds.

CPTIME_i = the CPU time during the ith stage of processing.

TIMEIO_i = I/O time during the ith stage of processing.

CM_i = number of central memory terms (octal) required during the ith stage of processing.

n = number of stages of processing.

a_1 = CPTIME cost percentage ratio per second = .3598/sec.

a_2 = TIMEIO cost percentage ratio per second = .5418/sec

a_3 = CM percentage cost ratio per 100,000 (octal) word second = .3936/100K word sec.

Organizational Utility. Jobs may be categorized in terms of the user's organization in order to evaluate the resource demands by each user. To find the proportion of resource usage by each organization, the number of jobs and the CRU's consumed by each organization are recorded.

Utility of Largest Programs. In many cases, a small subset of jobs may have a disproportionate effect upon the system; less than five per cent of the jobs often use more than thirty per cent of the system resources (Ref 4:24). In order to identify resource usage by larger programs in terms of CRU's consumed, the largest ten per cent of the jobs

processed during each investigation period are examined.

Hellerman's Objective Function. One approach to measuring the performance of a time sharing system involves the use of an objective function, or figure of merit. Hellerman defines a figure of merit as a function that reflects the intended use of the computer system (Ref 17:105). An example of such a function would be a measure of a system's ability to give higher priority service to requests for short jobs versus requests for long jobs. Hellerman's figure of merit includes the following four properties:

(1) It is dimensionless.

(2) It has a maximum value of one for some ideal scheduler and system for all workloads.

(3) Its value increases with system optimality; that is, the highest possible value of one would be attained when the execution time equaled the elapsed time. Values of less than one would occur when the elapsed time was greater than the execution time. The case of f=0 occurs when the elapsed time is infinitely large in that the job never completes execution.

(4) For a system with a job stream containing a mixture of both short and long jobs, Hellerman's figure is higher when the system provides better service to short job requests than to long job requests.

Hellerman's formula for the figure of merit which satisfies the above conditions is as follows:

$f = n / \sum_{i=1}^{n} (e_i / x_i)$ where

x_i is execution time for job request i.

$e_i = q_i - a_i$ = elapsed time for job request i.

a_i = the arrival time for job request i (the time it is first considered for execution by the system).

q_i = the time that a job request i completes execution.

n = the number of jobs investigated dur ng a time interval.

The figure of merit is intended to be a measure of a system's ability to give higher priority service to short versus long job requests. However, when used to analyze the multiprogrammed CYBER 74, it may be argued that the figure of merit also measures the degree of swapping over a time interval. The reason for this assessment is that in this study, the time a job enters a control point until it leaves a control point, is the value used for the elapsed time for a job request (e_i). Since the level of observation is limited by the available data, the elapsed time must include the time a job spends swapped out.

Data Collection System

A software program, Dayfile, which is maintained during normal processing by the NOS/BE operating system collects the accounting information for the CYBER 74 computer system resources. This program is primarily used to charge users for computer resources, but the measurements obtained may also be helpful in assessing computer system performance and workload measurement. The Dayfile data are saved on magnetic tapes and stored on a regular basis as a record of each day's activities. Unlike most software monitors which are active for only short periods of time, the Dayfile program is executed continuously by the operating system (Ref 29:2).

The accounting data may be divided into the three following categories (Ref 29:2):

(1) Identification data which includes the job name, name of person submitting the job, origin of the job and account number.

(2) Computer system resource requests and usage which include CPU time, PPU time, number of tape drives requested and used, I/O time, CRU's generated, and roll out time.

(3) Initiation and termination time which include time of day a job entered and exited a control point, and time a job entered the input queue.

CLARA. The Computer Load and Resource Analysis (CLARA) System, as discussed earlier, is a software program which extracts specific information from the Dayfile tapes. CLARA is discussed in more detail below.

CLARA reduces the large volume of Dayfile data to a much smaller data base. These data are stored on magnetic tape and may be processed to obtain additional information. The CLARA magnetic tape is more readily available and less expensive to use than is the Dayfile tape. Data from the CLARA tape are well suited for use in constructing a workload characterization and in evaluating the performance of the CDC CYBER 74 in that a wide range of parameters are available.

In order to be valid, the CLARA data must be analyzed carefully so that such problems as large numbers of dead starts and erroneous data are eliminated to avoid distorting the results. The time period investigated must be long enough to take into consideration the stability of the workload and operating system. Boeing Computer Service recommends a statistical base of about one month in order that the Dayfile data provide an adequate representation of system and job characteristics. The days and weeks of the time period under investigation may be analyzed individually once a base of sufficient time length has been obtained.

The CLARA program itself generates the following files for each day's Dayfile data:

(1) The Summary file which is a general description of the day's activities.

(2) The Tape-Reel file for the magnetic tape activities.

(3) The Queue file for information describing entry into the input queue.

(4) The Execution file which contains the majority of job parameters which describe the identification of each job, usage of resources, and timing of activities.

III. Program Implementation

This chapter presents the methodology used to characterize the accounting data. Due to the sheer bulk of this data, the machine-readable CLARA magnetic tapes were processed by the CYBER 74 computer system. Automatic processing has the following advantages (Ref 24:41):

(1) It permits easy repetition of the analysis over a time period involving several days of data.

(2) It permits analysis of several measurements of performance and workload.

(3) It eliminates most of the human error which occurs in processing large amounts of data.

The approach to the characterization of the CLARA data involved three steps: (1) preprocessing the CLARA variables by means of an Executive program, (2) analyzing the validated parameters by using the Statistical Package for the Social Sciences (SPSS), and (3) analyzing the long term workload trends and selecting the specific days for investigation.

First, the Executive program will be discussed; this will be followed by a short reference to the SPSS programs.

Executive Program Orientation

In order to obtain meaningful measurements on the CYBER 74 computer system, it was necessary to reduce the CLARA data to a manageable size and to eliminate invalid data. To achieve this goal, a computer program, the Executive, was designed and coded.

The objectives of the Executive program were (1) to reduce the CLARA parameters to a more manageable size, (2) to eliminate incorrect data, (3) to categorize the data for use in computing several initial

statistical measures, such as frequency histograms, and (4) to store validated variables.

The structure of the Executive program is diagrammed in Figure 3. The program consists of six operational modules which are controlled by a main module, EXEC; these seven modules and their purposes are listed below:

Module	Purpose
EXEC	To control the Executive program and to document the data structure.
INITIAL	To initialize the working variables and working arrays.
READIN	To read valid CLARA parameters from CLARA Queue and Execution files.
ITOTAL	To categorize parameters for intercom jobs.
BTOTAL	To categorize parameters for batch jobs.
STOTAL	To reduce and categorize all user system jobs for storage on a file.
DISPLAY	To print the results of initial statistical analysis.

The general flow of the Executive program is discussed in order to communicate the methods which were used to accomplish the program objectives. The Executive program is self-documented by using comment cards in order to provide a more precise description of the algorithms used.

The initial categorization of the jobs is divided into two major groups, batch and interactive. The batch jobs can be further subdivided into overhead batch jobs submitted by the ASD Computer Center, and user batch jobs submitted by the user organization. Interactive jobs include both INTERCOM jobs and graphics jobs. Since graphics jobs are not readily identifiable from INTERCOM jobs, they will be grouped as

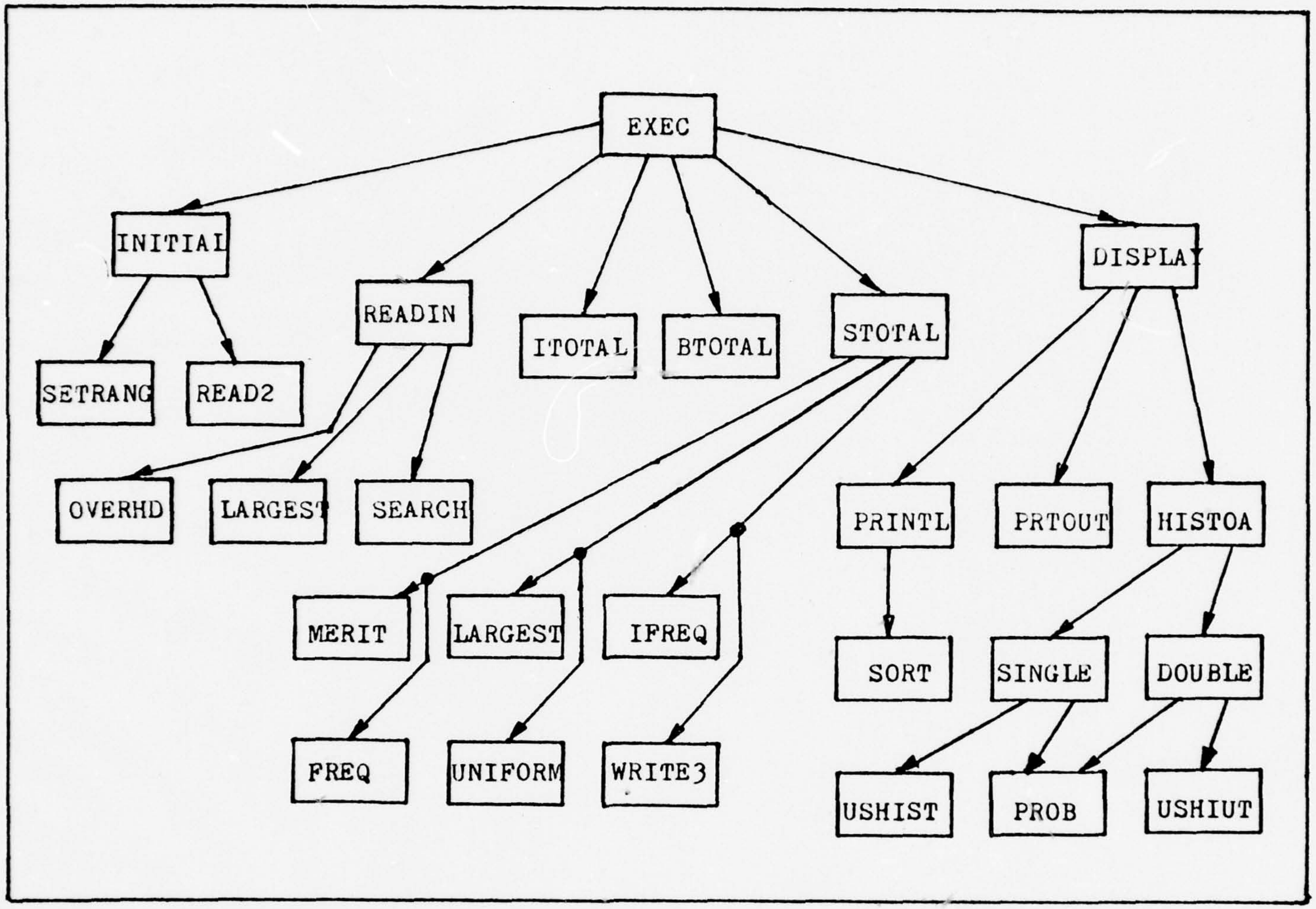


Fig. 3. General Structure of the Executive Program

interactive. This is not unrealistic because graphics jobs follow the same scheduling procedures as INTERCOM jobs.

EXEC Module. In order to preprocess the CLARA data, the EXEC module controls the six operational modules. INITIAL and DISPLAY are each called once during the processing of a CLARA file; INITIAL is called at the beginning, and DISPLAY is called after all the individual jobs have been processed. The four remaining modules, READIN, ITOTAL, BTOTAL, and STOTAL, are each called in sequence one time for each job that must be processed. ITOTAL is executed if the job is designated as a batch job.

INITIAL Module. The INITIAL module accomplishes five tasks prior to the execution of the remainder of the program:

(1) The first task is to ensure that erroneous data variables do not invalidate the calculations in the program; therefore, each working variable is initialized and each working array, which accumulates CLARA parameters for frequency distributions, is initialized.

(2) The scale values for categorization of the input variables into the frequency distribution histograms are accomplished by the module SETRANG.

(3) In order to establish the time boundaries of the investigation period, four parameters are read in from cards, or are set automatically by default. These parameters are INTER, the length of an interval; ITIME, the start of the investigation; NUMINTR, the number of intervals; and DATE, the date of the CLARA data file. Each CLARA file normally contains 24 hours of accounting data. A typical investigation period would begin at midnight and would run for 24 hours with intervals of one hour each. However, an investigation period could also be set to analyze an

eight hour period beginning at noon and having 24 intervals, each of 20 minutes duration.

(4) Since the input queue time (INPUTQ) is not available from the Execution file, another task of the INITIAL module is to gather the queue file data. The CLARA queue file is read by calling the READ2 module. The following four variables are retained for each job: the job name (JOB), the time a job entered into the input queue (INPUTQ), the time a job entered into a control point (INCP), and the time a job exited from a control point (OUTCP). The use of these variables enables the INPUTQ of the Queue file to be matched accurately with the correct job in the Execution file.

(5) The last task of the INITIAL subroutine involves setting the Execution file to the correct starting location for subsequent use by the READIN module.

READIN. The general function of the READIN module is to read one record containing an individual job's parameters from the CLARA Execution file each time the module is called, and to prepare the CLARA input parameters before they are processed further by other modules. In addition, there are several specific activities which were designed to validate the input variables.

(1) The range of initiation and termination times is checked and adjusted if necessary. For example, the time of day that a job enters an input queue (INPUTQ) is measured in seconds after midnight. Its range of values should be from zero seconds to a maximum of 86400 seconds (24 hours). Generally, the Dayfile accounting system is reinitialized every 24 hours at midnight; however, occasionally the clock used to record the accounting data is not reset to zero. As a result, some CLARA

times must be rescaled.

(2) In order to correctly categorize the job data for frequency analysis, the time interval that a job enters a control point and exits a control point is calculated.

(3) The memory term, kilo-word seconds (KWS), is calculated from the central memory term (CMTERM) by using a constant from the costing algorithm.

(4) Jobs outside the period of investigation are eliminated.

(5) Cumulative totals for all jobs identified as overhead are stored by calling module OVERHD. In addition, a special call is made to the LARGEST module to record the resource usage, in terms of CRUs, for the largest programs.

(6) All jobs with specific invalid parameters are eliminated.

(7) If a program is identified as a batch job, the module SEARCH is called to try to match the job names from the Queue file and Execution file in order to match the correct INPUTQ time from the queue file with the rest of the job data in the Execution file. Interactive jobs do not have an input queue time.

(8) When the last job is processed in the CLARA file, the DISPLAY module is called by the main program.

<u>ITOTAL and BTOTAL</u>. The function of the ITOTAL (BTOTAL) module is to formulate the frequency distributions of those jobs identified as interactive (batch) jobs. The number of jobs and amount of CRU's are separated into ten arrays representing the ten organizations placing demands upon the CYBER 74 computer system. The arrays will later be used to specify organization demands upon the system.

STOTAL. The STOTAL subroutine accomplishes a variety of functions designed to accumulate workload and performance parameters for all categories of jobs. Before listing the general tasks of the STOTAL module, a particular problem encountered during analysis of the accounting data will be discussed.

One disadvantage of an approach which uses accounting data exclusively, is that the computer resources used by each job are reported when the job is completed. This report would indicate that all the resources were actually consumed at the time of the job's completion. In cases where a large number of jobs are completed during a relatively short period of time, the measurement of the workload parameters would give an unrepresentative description of the workload and the performance during that time period. For example, during an hour interval, it is possible that the CPU time could indicate more than 60 minutes of active time.

A variety of techniques may be employed to more adequately represent the resource usage during an investigation period. Since the accounting data does not offer a means of calculating the actual distribution of resource usage, the following assumption was made:

> The resources consumed during the time a job holds a control point are evenly distributed over the entire period.

An example of the assumption can be applied to an investigation period with intervals of one hour. A job executing from 0930 to 1130 would have the system resources, such as the CPU, PPU, and I/O time, uniformly distributed: 1/4 of the total time accredited to the 0900 to 1000 period, 1/2 to the 1000 to 1100 period, and 1/4 of the total time distributed to the 1100 to 1200 interval.

The above assumption does not completely solve the problem of accurately distributing the accounting data, but it does give a more realistic representation of actual use.

The general tasks of the STOTAL subroutine are as follows:

(1) The first task of subroutine STOTAL is to increment the number of total jobs executed during the Executive program investigation period.

(2) The figure of merit is calculated by calling the subroutine MERIT.

(3) The names and computer resources consumed are saved for the largest jobs, in terms of CRU's, by calling the LARGEST subroutine.

(4) Each of the workload parameters is processed next. First, subroutine UNIFORM is called, if applicable, to distribute the resources into the correct time intervals. The parameter quantities are accumulated in the array location corresponding to time interval for later frequency analysis. Finally, IFREQ (for integer variables) or FREQ (for real variables) module are called to increment the arrays used to monitor the frequency distribution of each variable.

(5) The validated workload and performance parameters are then written to a file by the WRITE3 module.

DISPLAY. The function of the DISPLAY module is to print on the line printer, the results of the frequency analysis, probability functions, and several performance measures. The general tasks of the module are as follows:

(1) The amount of resources consumed by the largest 10% of the programs and by the overhead jobs are printed by calling the PRINTL module. In addition, after calling the SORT module, the largest jobs

are listed in order by size.

(2) The arrays containing the frequencies of occurrence for each parameter are listed by calling the PRTOUT subroutine.

(3) The final task of DISPLAY is to print the frequency histograms of distribution for each variable and frequencies of occurrence during the investigation period. The task is aided by SINGLE and DOUBLE subroutine for single and double histograms. These subroutines also call PROB for the probability distribution functions and two utility subroutines: USHIST and USHIUT for the actual histogram representations.

Further investigation of the preprocessed CLARA data is completed by another utility program, the Statistical Package for the Social Sciences, which is discussed in the following section.

Statistical Package for the Social Sciences

The Statistical Package for the Social Sciences, a validated and tested computer program, was selected to calculate the workload statistics necessary for this thesis. The SPSS system of computer programs represents more than ten years of designing and programming, and has been valuable to social scientists, statisticians, and computer specialists (Ref 21:xxi). Use of this package offers the advantage of reducing the amount of coding effort needed to process the CLARA parameters.

The basic statistics on the workload characterization will be completed by the program CONDESCRIPTIVE. Further information on this program may be gathered from Reference 21.

IV. Results

This chapter presents the results of the workload investigation upon the CYBER 74 accounting data. The discussion is presented in the following order:

CYBER Workload Trend Analysis

Problems Encountered with Accounting Data

Workload Characteristics

Performance Measurement

CYBER Workload Trend Analysis

The available CLARA accounting data encompasses more than six months of background information which is recorded on magnetic tape. This historical data was the base for an investigation of workload and performance trends on both a long term (days) and short term (hours) basis. The results of the trend data were then used to select a time period and specific days for further investigation.

Four performance measures were examined for a six month investigation period which included normal weekdays only. These performance measures, total jobs executed per day, CPU utilization per day, machine turnaround time per day, and average CRU's per hour for each day were selected because they were readily accessible and were representative of the available accounting data. Weekdays were selected since a majority of the complaints about slow response time and turnaround time occurred then, and because the CLARA data on weekdays were significantly different from the weekend workload; specifically, the weekend workloads were dominated by a small number of jobs (less than 500) with long execution times.

The observations below were made from the following four figures; Table 3 lists the days which were used in the trend analysis.

(1) The total jobs executed per day (Fig 4) ranged from 850 to 1700. Since the INTERCOM jobs were relatively constant, the total number appeared to be a function of the number of batch jobs.

(2) The CPU utilization (Fig 5) showed an overall increasing trend which began on the 37th day (23 June 1976) of the investigation and continued until the 95th day (15 September 1976). Utilization ranged from 41 per cent to 88 per cent.

(3) The machine turnaround time, the time from job entry into the input queue to the final exit from the control point (Fig 6), varied from a daily average of 0.25 (9 July) to 1.70 hours per job (7 October).

TABLE III

Trend Analysis Days and Dates

Days	Dates	Days	Dates
1- 5	1 to 5 May	64-68	2 to 6 Aug
6-10	10 to 14	69-73	9 to 13
11-15	17 to 21	74-78	16 to 20
16-20	24 to 28	79-83	23 to 27
21-24	1 to 4 Jun	84-85	30 to 31
25-29	7 to 11	86-88	1 to 3 Sep
30-34	14 to 18	89-92	7 to 10
35-39	21 to 25	93-97	13 to 17
40-42	28 to 30	98-102	20 to 24
43-44	1 to 2 Jul	103-106	27 to 30
45-48	6 to 9	107	1 Oct
49-53	12 to 16	108-112	4 to 8
54-58	19 to 23	113-116	12 to 15
59-63	26 to 30	117-121	18 to 22
		122-126	25 to 29

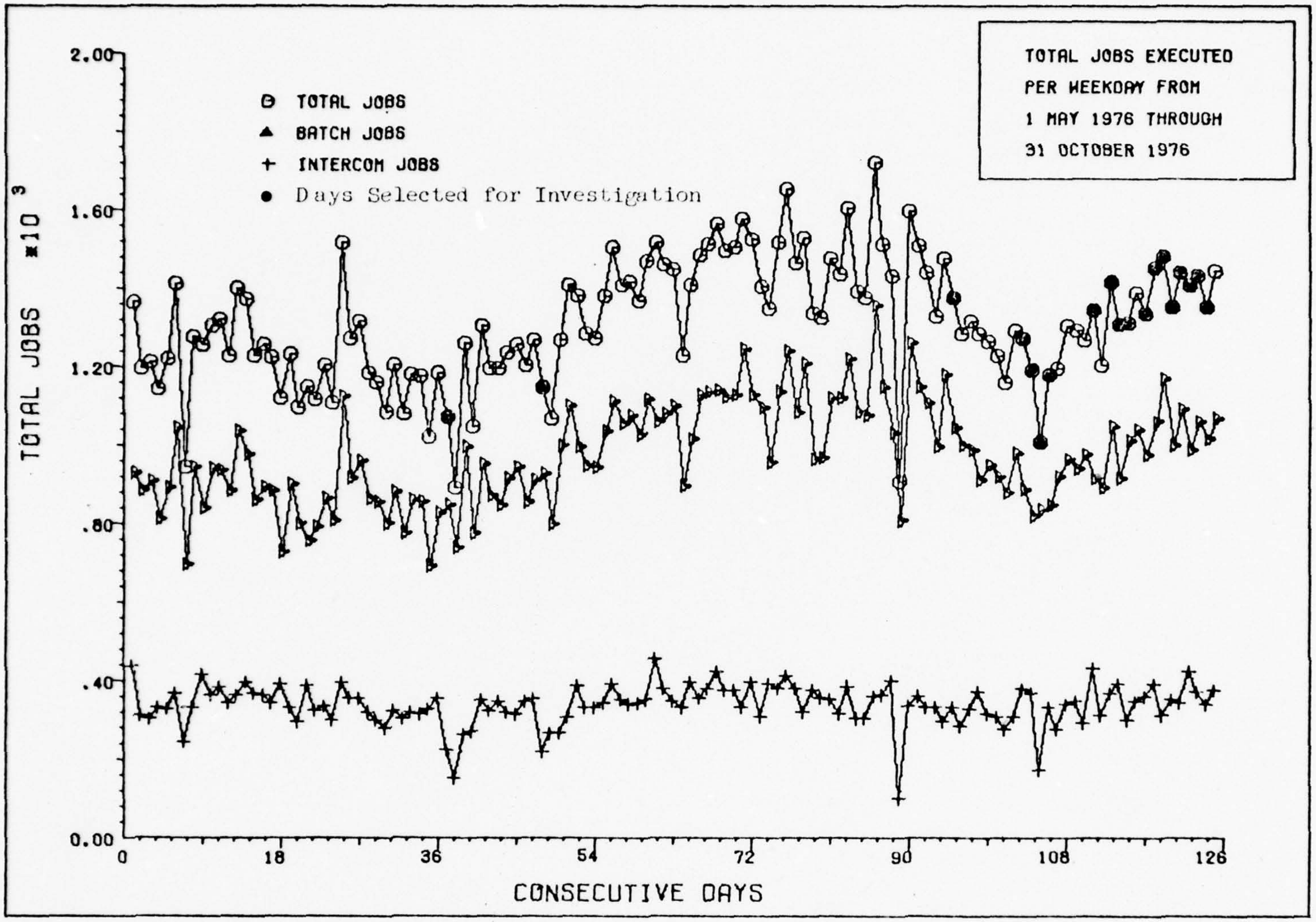


Fig. 4. Trend Analysis for Total Jobs per Day

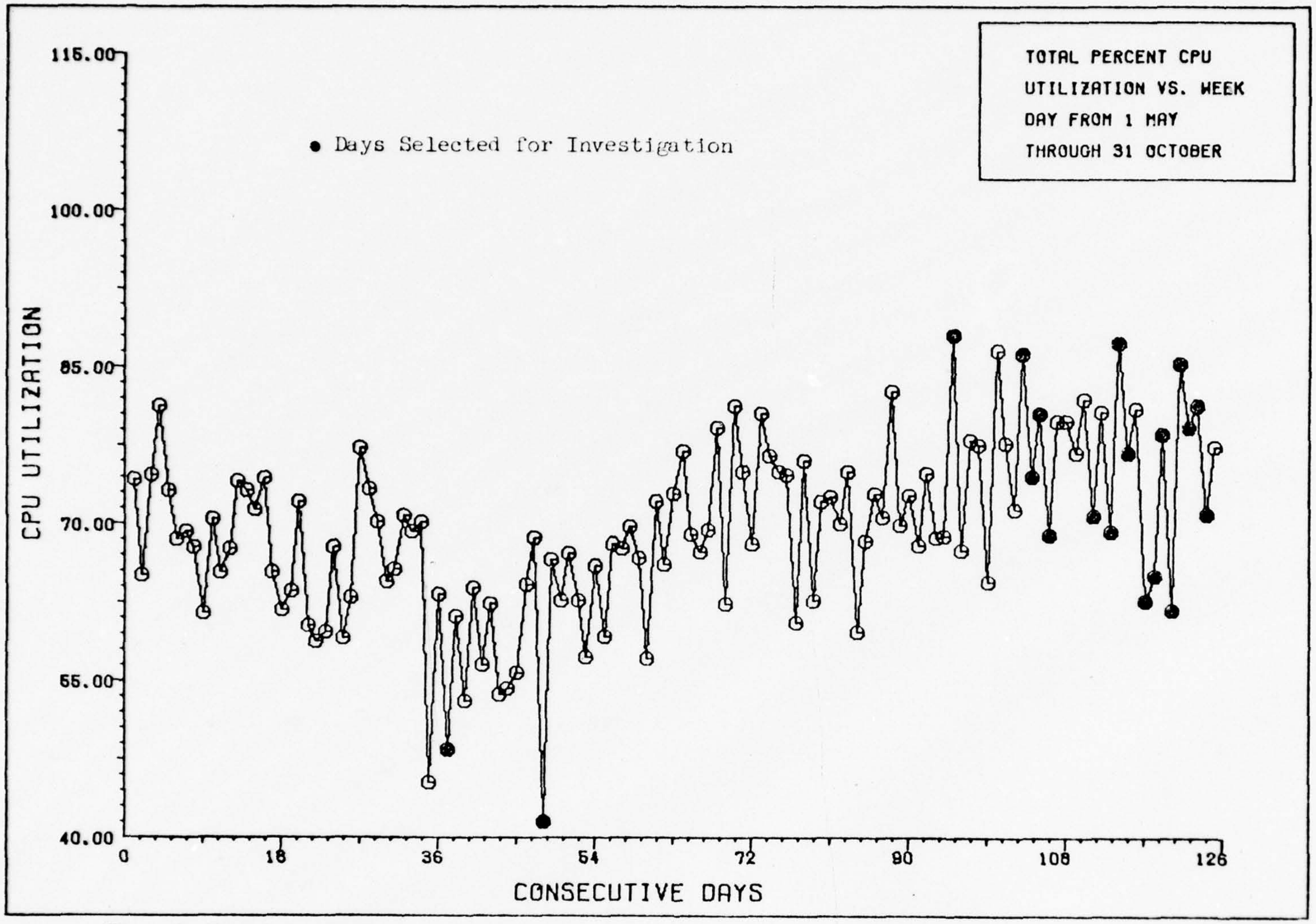


Fig. 5. Trend Analysis for CPU Utilization

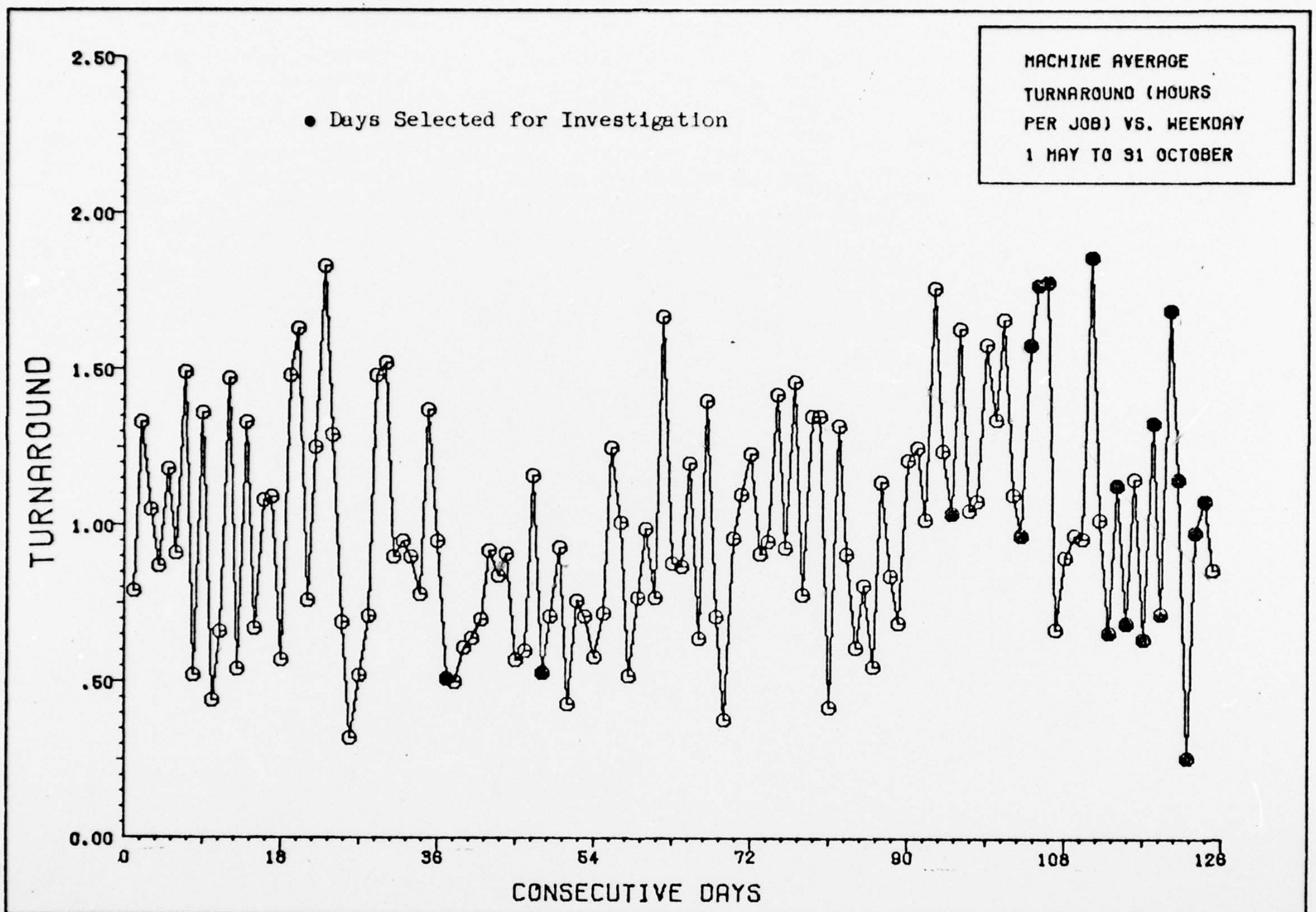


Fig. 6. Trend Analysis for Turnaround

(4) The average CRU's per active hour (Fig 4) ranged from zero to 1200 CRU's per hour for INTERCOM jobs. Batch jobs, more variable in production, ranged from 2600 to 5500 CRU's per hour. In order to normalize the measure, CRU's produced per hour were measured rather than CRU's produced per day.

Further, trend analysis over longer periods of time may also confirm weekly, monthly, and quarterly cycles; the trend data appears to be periodic. For example, the workload upon the computer system appears to be heavier at the beginning of a week. On a quarterly basis, the number of jobs tends to increase and decrease as the (academic) quarter begins and ends at the Air Force Institute of Technology (Fig 7). An investigation into these longer term phenomena was not undertaken due to the time limitations upon this thesis.

Several days were selected for further investigation as seen in Table IV. The number of days was limited to 21 due to the sheer amount of data to be processed. The days were selected to be representative of average performance as well as extremes in performance. In order to minimize the effects of changes in the computer system, such as changes in the operating procedures, computer center personnel, and general demand for computer services, the days were generally grouped within a 30 day period. The investigation period ranged from the 103rd day (27 September 1976) through the 125th day (28 October 1976) in order that the most recent workload data be investigated. Three other days which were distinguished by their uniqueness, were also included in the investigation period. These were 23 June, 9 July, and 15 September.

It should be noted that the variation in the INTERCOM CRUS generated from the 96th through the 125th day in Figure 7 was caused by a change in the accounting method for this parameter rather than by a change in the performance measurement of the CRU's per hour.

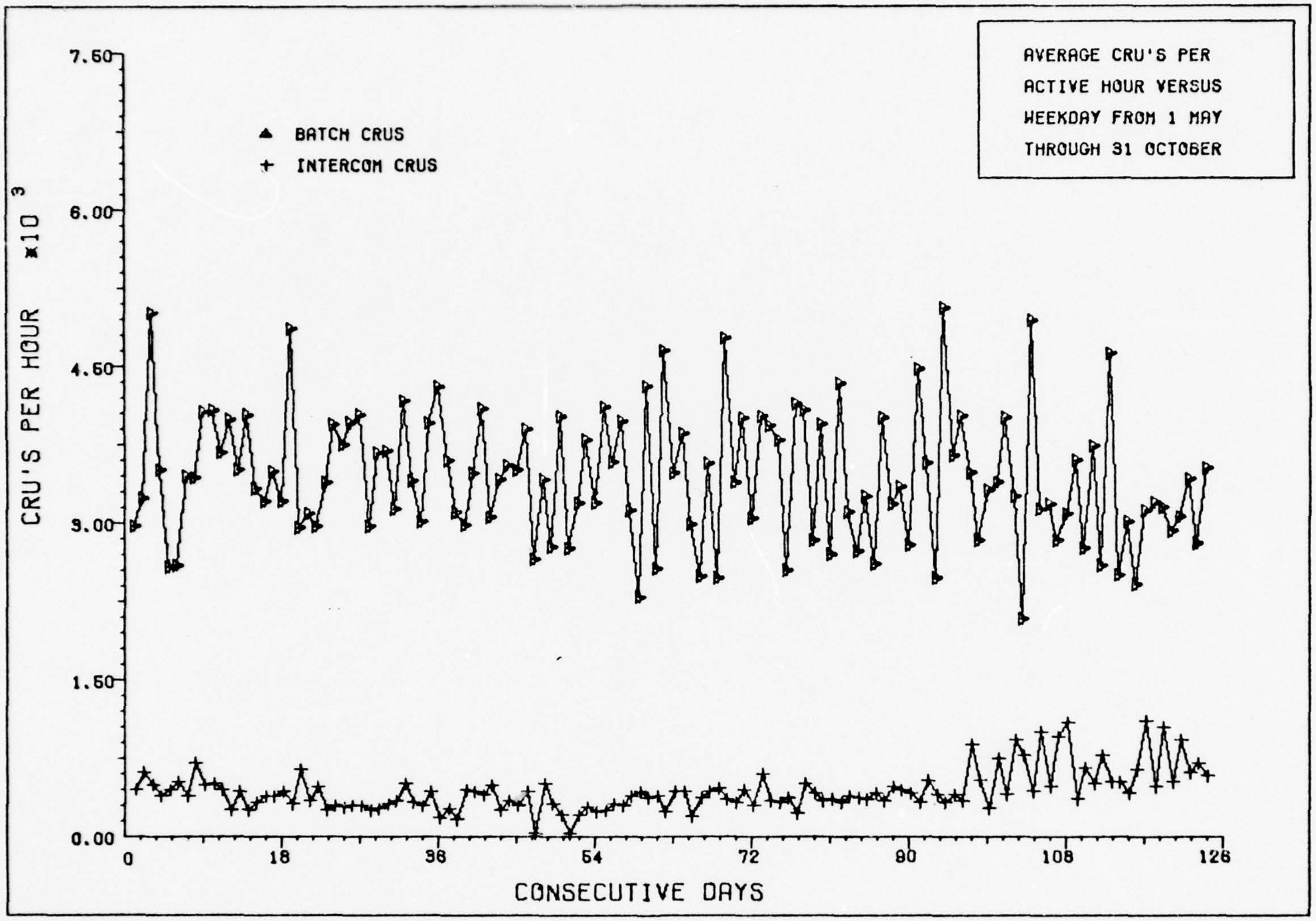


Fig. 7. Trend Analysis for CRUs per Job

TABLE IV

Days Scheduled for Investigation

Day	Date	Day	Date
1	23 June*	11	13 October**
2	9 July*	12	14 October
3	15 September**	13	18 October**
4	27 September**	14	19 October
5	28 September	15	20 October
6	29 September**	16	21 October
7	30 September	17	22 October*
8	7 October	18	25 October**
9	11 October**	19	26 October
10	12 October	20	27 October**
		21	28 October*

*Excluded from further investigation as non-representative.

**Reduced sample of data for INPUTQ parameter; thus, the following parameters are calculated from a reduced data base: time in input queue, turnaround time, figure of merit, interarrival time.

In order to eliminate nonrepresentative data, the criterion was established that a day would not be included in the investigation if malfunctions of the central computer site resulted in idle time for that day which comprised four per cent or greater of total active time; two of the 21 days investigated, 23 June and 9 July, were excluded on this basis. Two other days, 22 and 28 October, were also excluded due to the extremely low number of jobs recorded on the CLARA tape. The remaining 17 days were used as a data base for this investigation.

During the period of exploratory investigation, the entire 24 hour day was studied; however, it was necessary to narrow this time period because the maintenance of the computer system, which is periodically scheduled at midnight for several hours, was biasing the results

obtained. For example, the number of CRU's generated per day varied widely depending upon the number of active hours. A second approach was attempted which focused upon the 0800 to 2400 time period. Several problems were encountered, and it was decided to reduce the period of study to a more specific workload, the hours from 0800 to 1600, for the following reasons:

(1) The largest number of users are active on the CYBER 74 during normal working hours from 0800 to 1600.

(2) Since the INTERCOM activity is relatively constant during the day and tends to be reduced in the evening, the daytime period is more appropriate for measurement of the CYBER 74 multiprogramming mode, including batch and intercom.

(3) In an attempt to give the smaller user a faster response and turnaround time during the day, the ASD central site withholds until 1600 each day, those jobs requiring a large number of system resources.

(4) The majority of complaints from users occur during the 0800 to 1600 time period, and relate to slow turnaround and response time.

Unless stated otherwise, the remaining discussion of the results will refer to the 0800 to 1600 time period for the 17 investigation days.

Problems Encountered with Accounting Data

Some deficiencies were discovered in the information obtained from the CLARA data. In order to place the later discussion of the workload characterization and performance measures in proper perspective, and in order to provide a background for further investigations, the problems involving inadequate information in the accounting data will be discussed first.

On the whole, the CLARA parameters provided most of the information necessary for a relatively complete investigation of the workload and performance. However, since some of the parameters recorded on the

Execution file of the CLARA magnetic tapes proved to be deficient, it is necessary to discuss how these information problems affected workload and performance measures, and how these problems were dealt with. The problem involves parameters which were either not produced for storage on the CLARA magnetic tape, or parameters which were found to be misleading or incorrect. For future investigation, it may be possible to extract the correct data from the original Dayfile accounting tapes. The discussion which follows will include some variables which were not used for the study because of missing or invalid data.

(1) The most important problem involved the time of day that a job entered the input queue (INPUTQ). Since the INPUTQ was not available on the Execution file, it was necessary for the Executive program to read and store from the Queue file, the job names, INPUTQ times, and times of entry (INCP) and exit (OUTCP) from a control point. As the jobs were read from the Execution file, the corresponding INPUTQ was matched to the job by use of the SEARCH subroutine. In addition to computer run time being tripled for processing a day of CLARA data, because of the INPUTQ not being available on the Execution file, another problem was encountered.

Further investigation of the INPUTQ times recorded on the Queue file revealed that often up to 50% of the jobs were not saved on the Queue file. Since some parameters were derived from INPUTQ times, it was decided that the data from any day having a Queue file with less than 96% job representation would be identified, as seen in Table IV. Depending upon the specific day investigated, the following parameters

may be based upon a reduced data base.

(a) Interarrival time (IATIME), the time between jobs arriving into the input queue.

(b) Average time spent by a job in the input queue (average INQTIME).

(c) Figure of Merit, used to measure the responsiveness of the computer system to the small job.

(d) Turnaround time, the time a batch job arrives into the input queue until it leaves a control point.

(2) Other variables which were stored on the Execution file could not be used because they were either incorrect or misleading. For example, the job card field length (CMREQ) was always recorded as 5000 regardless of the actual request on individual job cards for central memory use. Another parameter, the maximum central memory required by the controller (CMCOMP), was unusable because it was always recorded as zero on the Execution file; as a result, it was impossible to judge whether there actually had been no use of the central memory by the compiler or whether the memory had been used but had not been recorded.

(3) The total magnetic drive tape time (X1) was not used in this thesis because it was always listed as one (1) second. In order to overcome this problem, a substitute parameter was developed which correctly reflected the amount of magnetic tape time used. This parameter, called ROLLOUT, could be calculated for each job by adding two variables, TOTROLL and LSTROLL, from the Execution file. TOTROLL is the total time a job is rolled out to process magnetic tapes (except for the last tape). LSTROLL is the total time a job was last rolled out.

(4) Since the amount of memory used by each job was not directly available, it was necessary to recalculate the kiloword seconds of memory (KWS) from the central memory term (CMTERM) which was listed on the Execution file.

(5) Since an information entity was not listed regarding the organization submitting the job or whether the job was interactive, batch, or overhead, an attempt was made to derive this data from the job name (JOB) and the account number (ACCTNUM). Unfortunately, ACCTNUM was often missing from data for batch jobs; as a result, any batch job submitted at the Computer Center could not be distinguished by organization. The results of this problem are discussed later in the performance results.

(6) Another problem confronted in using the input queue time was the fact that although in some cases the SEARCH subroutine matched the job name from the Execution file to the job name in the Queue file, the INCPs and OUTCPs of both files did not correspond exactly. An algorithm in the SEARCH subroutine determined that less than 1% of the jobs were affected; thus, the elimination of this data has little effect upon the overall results.

(7) A final problem encountered in using the Queue file information was that the INPUTQ was occasionally recorded as occurring later than the INCP, although this is not a design possiblity. Also, the INCP and OUTCP for the same jobs, particularly overhead jobs, were recorded with the result that the time spent by a job in the control point (CPCT = OUTCP-INCP) was less than the amount of central processor time (CPTIME) used during the job; this was also a design impossibility. Since less than one (1) per cent of the jobs were affected, the

elimination of this invalid job data should have only a minor impact upon this investigation.

Table V contains a list of variables on the CLARA Execution file which was investigated and not used; following each variable is the reason(s) for exclusion from the thesis study.

Workload Characteristics

For the investigation days, the accounting data was characterized for each day by statistically measuring 13 workload parameters. Workload characterizations obtained from several representative days are summarized in Tables VI through VIII. In addition, several workload characterizations were separated into types of job classes, those based upon CPU time, I/O time, and memory used. Each of these classes consists of a workload characterization containing low, medium, and high values of the workload parameter upon which the respective class is based. The six days presented in Table VI comprise the two days which averaged the lowest CPU time per job, the two days which average the highest CPU time per job, and two days in which the CPU average per job was in the middle range.

In addition, selected workloads are more fully characterized in the appendices by a frequency histogram and a probability density function for each parameter. A statistical summary for each of the workload parameters may also be found in each appendix. The information in the appendices could be used in a future investigation to develop a workload for a performance evaluation.

TABLE V

Variables Excluded from Investigation

Name	Description	Reason
PERSON1,2	Name of person submitting job.	Outside scope/ Not available
JOBORGN	Location or origin of job.	Not available
PHONE	Phone number of person submitting the job.	Not available
JOBCODE	Job origin code of location of physical terminal.	Not available
CMREQ	Central memory requested (field length).	Not available
CMCOMP	Maximum central memory required by compiler.	Not available
CMLOAD	Maximum central memory required to load.	Not available
FLDLNG	Last field length appearing on a Request for Field Length card.	Not available
FILNAM	File name appearing on the last request control card.	Not available
X1	Total drive tape time.	Invalid
ERRNUM	Error number, if any, that occurred during processing.	Outside scope
LASTECS	Last Extended core storage (ECS) appearing on a request for ECS control card.	ECS not installed on CYBER 74
JOBECS	Job card ECS field length.	ECS not installed on CYBER 74
CMPCALL	Number of times the compiler was called.	Invalid
DPMS	Total accumulated length of all dumps.	Not available
PRGMS	System programs used.	Outside scope
JBPRIOR	Job card priority.	Not available
EQUIP	Equipment use in processing.	Outside scope
PRGUSED	Programs used by name.	Not available
LASTRFL	Time of last request for field length.	Invalid

CPU Time. The first set of job classes is based upon CPU time for INTERCOM and batch jobs. The summary of the workload characterizations is presented in Table VI.

It is interesting to note that the distributions for job CPU time are similar across the entire range of job classes. For example, the probability density functions for 29 September and 20 October, shown in Figures 8 and 9, indicates that over 50 per cent of the jobs used less than six seconds of CPU time. However, as may be seen in Table VI, the average CPU time for all job classes, is skewed higher since there is a large group of jobs (over 15 per cent), each of which uses more than 40 seconds of CPU time.

The parameters which comprise the workload for 20 October, 7 October, and 29 September are statistically characterized in Appendices A, B, and C, respectively.

I/O Time. The set of job classes based upon I/O time is presented in Table VII. The three specific job classes are based upon those days with low, middle, and high I/O times per job.

The wide range of average I/O time per job is readily apparent since the ratio between the largest average and the smallest is 100 times. Comparison of I/O time per job for several days indicated the high average I/O per job may be a result of the large number of big jobs being processed by the computer system. The difference between the high and low I/O job classes can be seen in Figures 10 and 11.

Appendices D, B, and E contain workload descriptions for 25 October, 7 October, and 30 September, respectively.

TABLE VI

CPU Workload Characterization

WORKLOAD PARAMETERS (AVG VALUE PER JOB)	11 OCT	20 OCT	7 OCT	21 OCT	29 SEP	27 SEP
CPTIME	21.67	22.13	29.31	29.68	35.17	35.53
PPTIME	38.36	48.45	98.03	72.02	95.72	56.62
TIMEIO	17.98	24.02	881.05	16.34	1397.72	808.17
TOTCOST	26.73	31.00	31.46	29.44	43.55	27.66
KWS	707.43	834.04	881.25	826.16	1397.78	816.16
CPOT	903.66	1080.86	804.03	964.20	1158.486	760.25
CMLOC	142858	187128	149777	146110	202640	166231
CPLOC	88.20	95.44	61.21	96.42	92.19	67.57
ROLLOUT	21.00	22.33	22.46	16.17	61.29	16.19
LATIME	47.28*	46.04	65.35	50.23	49.92*	44.88*
INQTIME	598.58*	1651.24	808.29	831.63	799.85*	322.38*
TAPEREQ	.060	.057	.059	.056	.078	.073
TAPEUSED	.053	.053	.055	.053	.076	.067

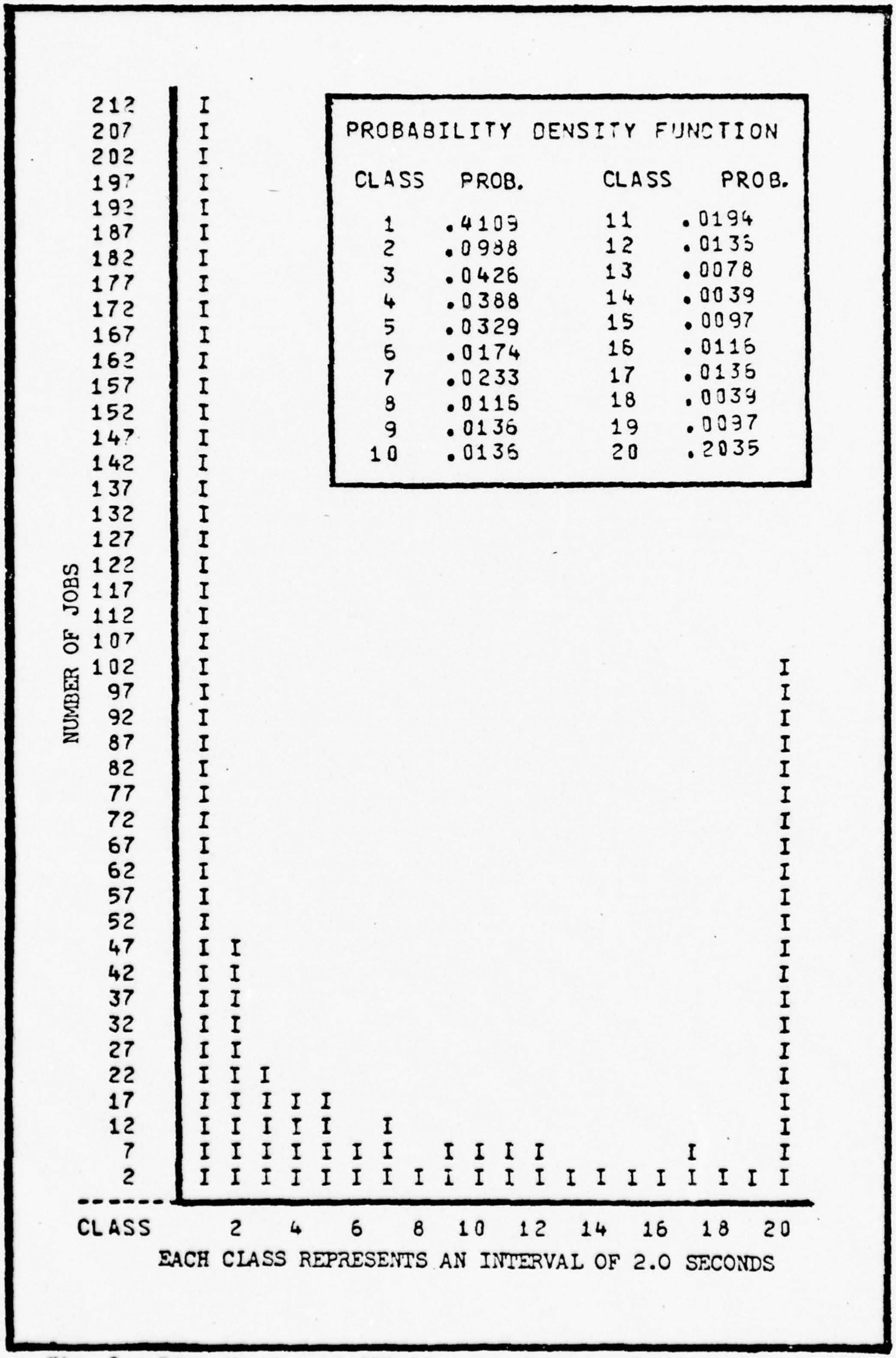
*Value based upon reduced data base for this parameter.

TABLE VII

IO Workload Characterization

WORKLOAD PARAMETERS AVG PARAMETER	25 OCT	21 OCT	28 SEP	7 OCT	30 SEP	29 SEP
CPTIME	24.28	29.68	37.43	27.31	29.15	35.17
PPTIME	78.36	72.02	66.50	98.03	83.76	95.72
TIMEIO	13.95	16.34	839.96	881.05	1056.61	1397.72
TOTCOST	24.08	29.442	32.49	31.46	36.52	43.55
KWS	648.56	825.16	839.98	881.25	1056.70	1397.78
CPOT	1045.16	964.20	1142.60	804.03	1103.85	1158.486
CMLOC	191314	146109	51228	149777	180053	202640
CPLOC	90.32	96.42	96.86	61.21	91.83	92.19
ROLLOUT	22.13	16.17	41.63	22.46	42.15	61.29
LATIME	40.06*	50.23	69.83	65.35	101.36	49.92*
INQTIME	129.74*	831.63	1218.00	808.29	1533.99	799.85*
TAPEREQ	.052	.056	.093	.059	.097	.078
TAPEUSED	.052	.053	.092	.055	.074	.076

*Value based on reduced data base for this parameter.



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.4109	11	.0194
2	.0988	12	.0135
3	.0426	13	.0078
4	.0388	14	.0039
5	.0329	15	.0097
6	.0174	16	.0116
7	.0233	17	.0136
8	.0116	18	.0039
9	.0136	19	.0097
10	.0136	20	.2035

Fig. 8. Distribution of CPU time for 29 September

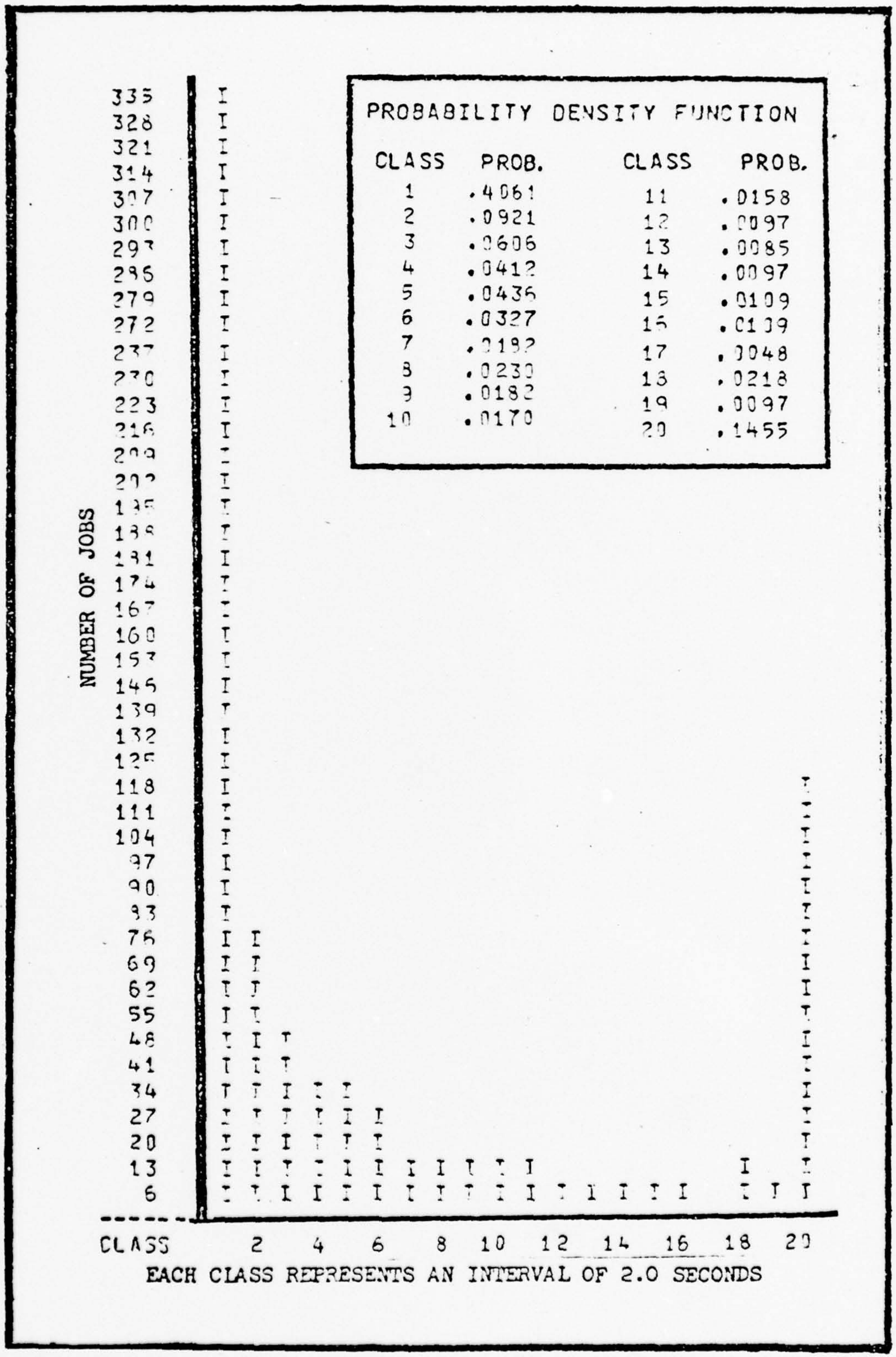
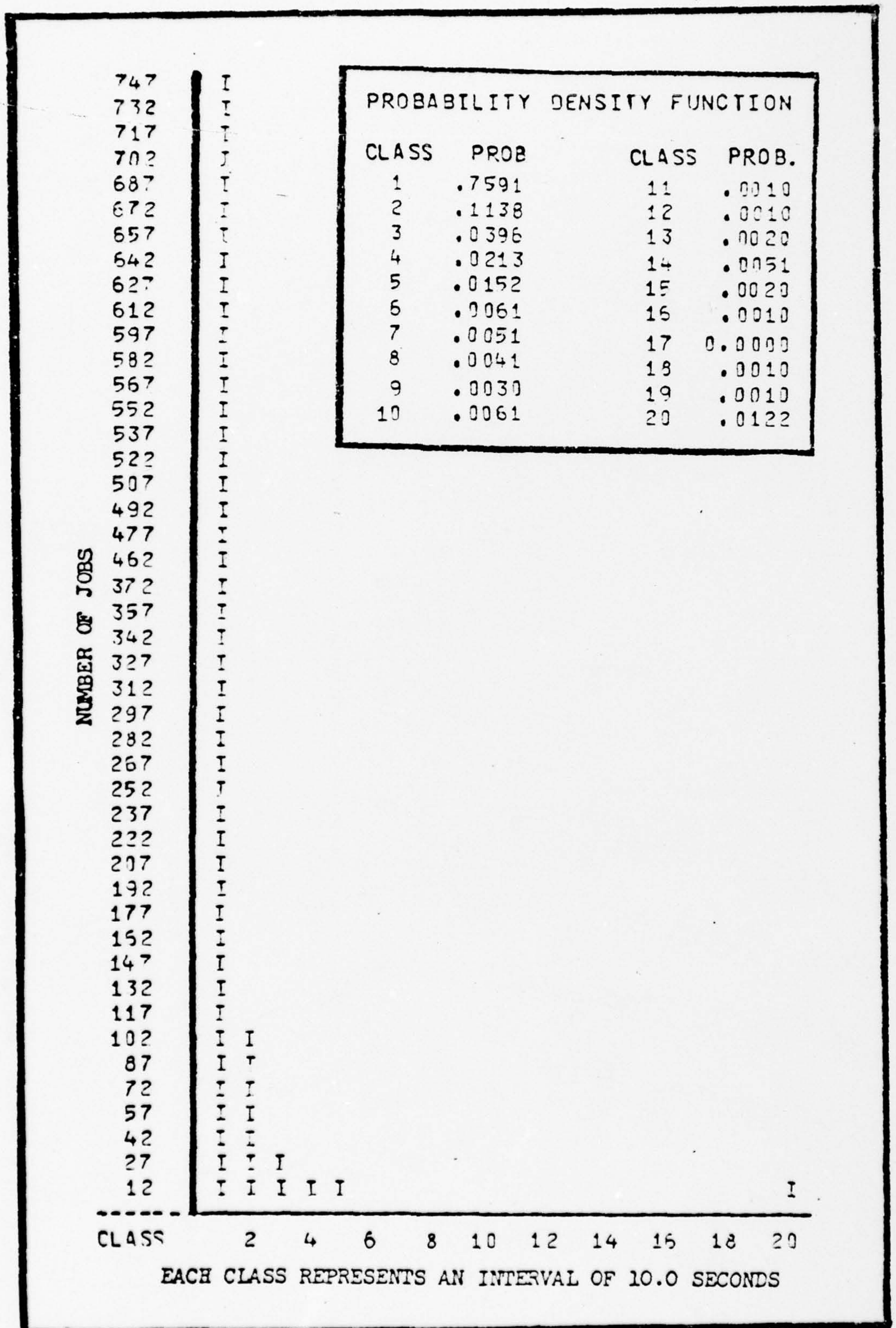
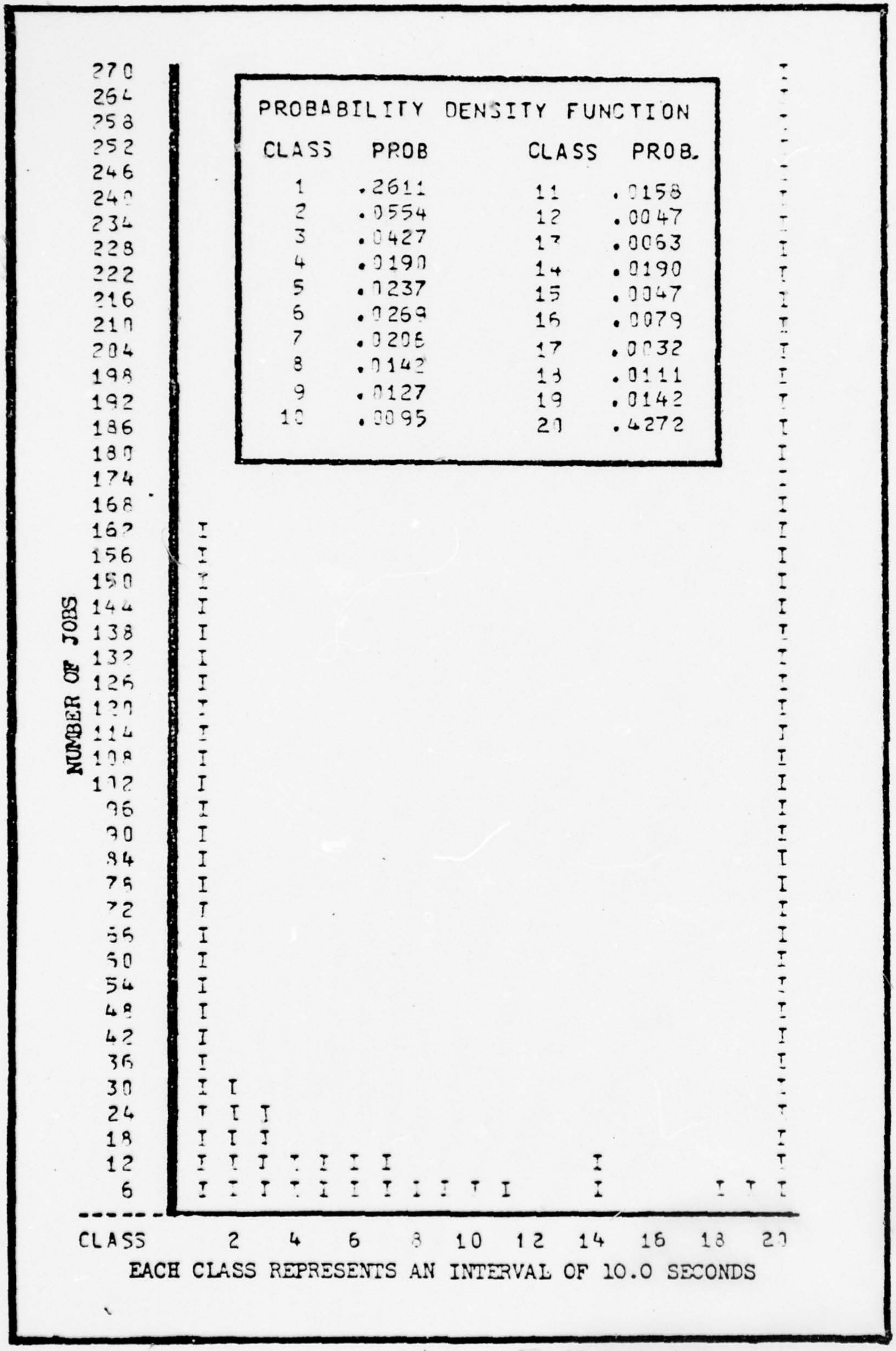


Fig. 9. Distribution of CPU time for 20 October



CLASS	PROB	CLASS	PROB.
1	.7591	11	.0010
2	.1138	12	.0010
3	.0396	13	.0020
4	.0213	14	.0051
5	.0152	15	.0020
6	.0061	16	.0010
7	.0051	17	0.0000
8	.0041	18	.0010
9	.0030	19	.0010
10	.0061	20	.0122

Fig. 10. Distribution of I/O for 25 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.2611	11	.0158
2	.0554	12	.0047
3	.0427	13	.0063
4	.0190	14	.0190
5	.0237	15	.0047
6	.0269	16	.0079
7	.0206	17	.0032
8	.0142	18	.0111
9	.0127	19	.0142
10	.0095	20	.4272

Fig. 11. Distribution of I/O time for 30 September

KWS. The last set of job classes is based upon memory used per job in terms of kiloword seconds (KWS). Depicted in Table VIII, the amount of memory used by the jobs also are characterized as being a lower average job cost (TOTCOST). For both high and low job classes, more than 50 per cent of all the jobs use less than 150 KWS as may be seen in Figures 12 and 13. Again, the amount of memory used appears to be a function of the number of large jobs being processed by the computer system.

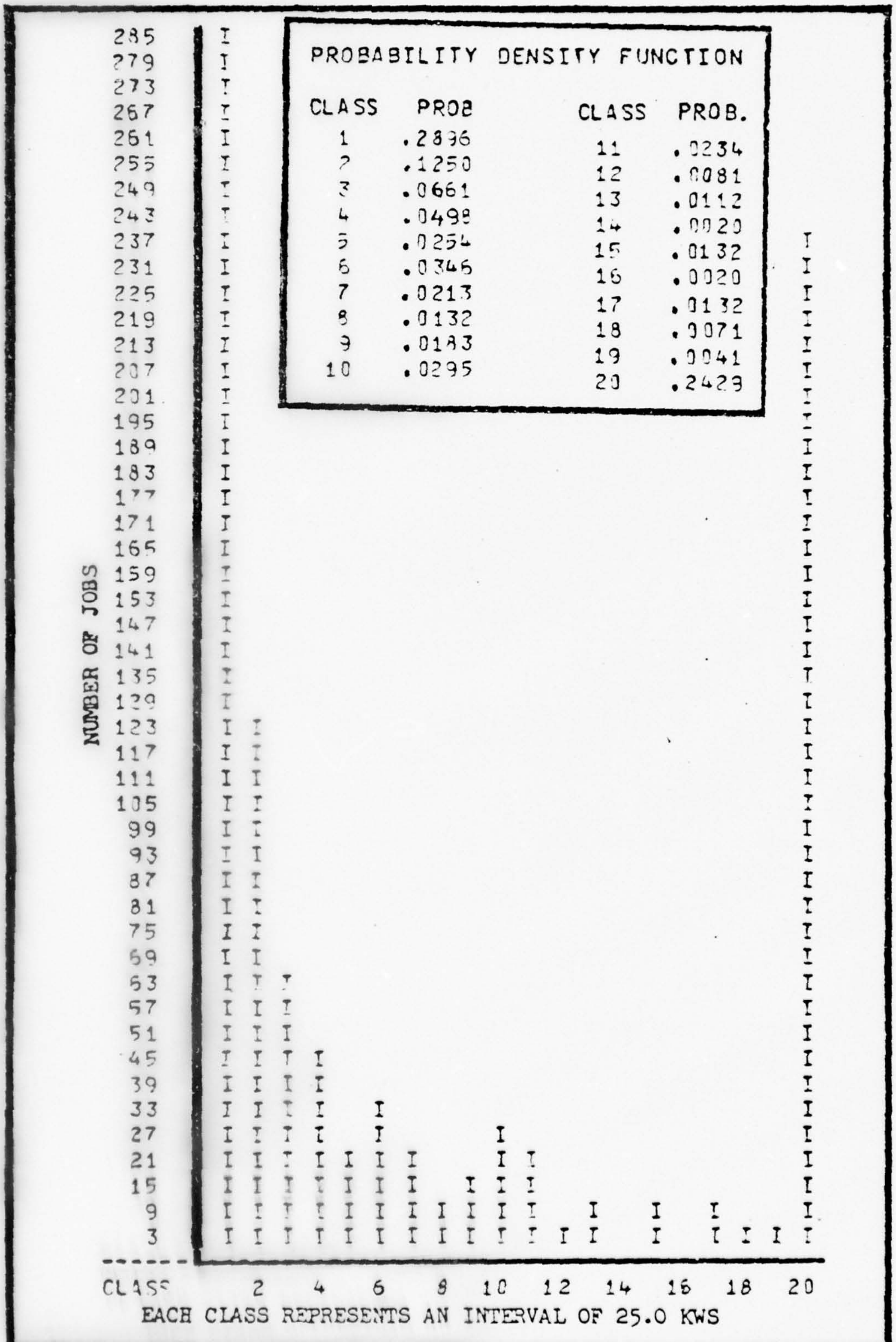
Appendices D, G, and H statistically characterize the workloads for 25 October, 19 October, and 13 October, respectively.

TABLE VIII

KWS Workload Characterization

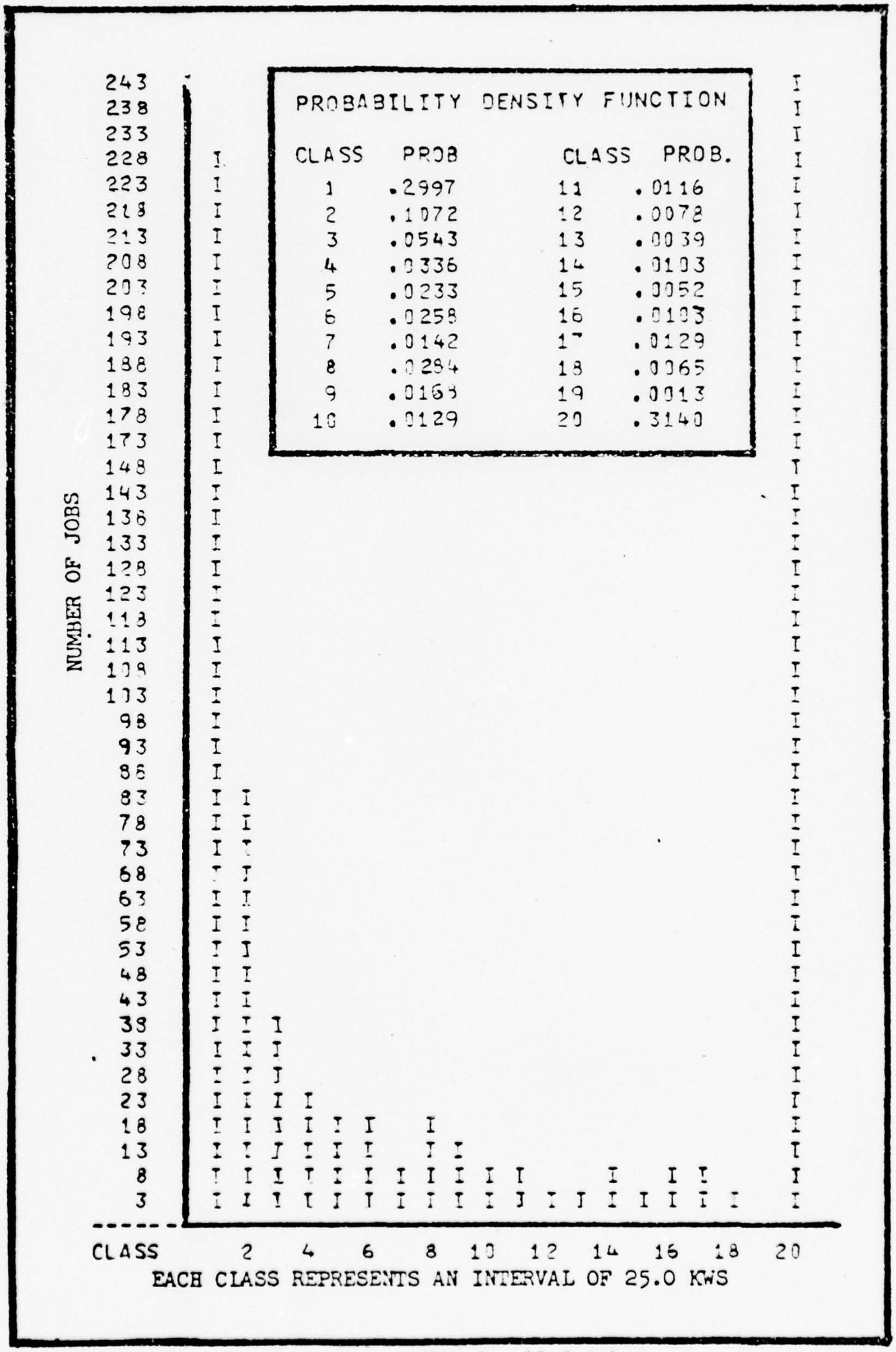
WORKLOAD PARAMETERS (AVG VALUE PER JOB)	25 OCT	14 OCT	19 OCT	27 OCT	29 SEP	13 OCT
CPTIME	24.28	22.60	23.12	23.26	35.17	34.10
PPTIME	78.36	62.36	34.53	68.21	95.72	73.83
TIMEIO	13.95	15.49	22.95	23.53	1397.72	17.13
TOTCOST	24.08	29.96	31.43	33.40	43.55	35.81
KWS	648.56	701.93	868.88	1022.81	1397.78	1187.30
CPOT	1045.16	777.24	863.50	1234.32	1158.486	1176.29
CMLOC	191314	124030	129792	76375	202640	129083
CPLOC	90.32	94.85	97.25	95.27	92.19	91.95
ROLLOUT	22.13	17.37	27.50	16.83	61.29	41.82
IATIME	40.06*	44.50	49.52	42.22*	49.92*	41.49*
INQTIME	129.74*	576.55	1310.99	645.79*	799.85*	345.60*
TAPEREQ	.052	.062	.048	.050	.078	.074
TAPEUSED	.052	.058	.047	.050	.076	.071

*Value based upon reduced data base for this parameter.



CLASS	PROB	CLASS	PROB.
1	.2896	11	.0234
2	.1250	12	.0081
3	.0661	13	.0112
4	.0498	14	.0020
5	.0254	15	.0132
6	.0346	16	.0020
7	.0213	17	.0132
8	.0132	18	.0071
9	.0183	19	.0041
10	.0295	20	.2429

Fig. 12. Distribution of KWS for 25 October



CLASS	PROB	CLASS	PROB.
1	.2997	11	.0116
2	.1072	12	.0078
3	.0543	13	.0039
4	.0336	14	.0103
5	.0233	15	.0052
6	.0258	16	.0103
7	.0142	17	.0129
8	.0284	18	.0065
9	.0168	19	.0013
10	.0129	20	.3140

Fig. 13. Distribution of KWS for 13 October

A general observation for all three job classes is that the workload for the CYBER 74 generally is distributed bimodally. One group, which comprises approximately 80 per cent of the jobs, is oriented toward large numbers of small jobs. The other group is composed of about 20 per cent of the jobs and consists of larger jobs which consume large amounts of CPU, I/O, or memory. The dominant factor in the computer system workload appears to be the number of large jobs being processed.

Performance Measurement

The effects of variations in workload upon the computer system performance have been measured by a variety of techniques as discussed in the Methodology chapter. Three measures, overhead job utilization, largest jobs utilization, and organization utilization, are first described in order to project an overall view of computer performance.

An analysis of the following performance measures was conducted each day for the time interval from 0800 to 1600:

CPU Utilization

Turnaround Time

Thruput Time

CRU Generation

Figure of Merit

Before discussing the performance measures, one anomaly should be noted: the distribution in the histograms for some parameters may be misleading. For example, in Figure 17 the CPU utilization from 0500 to 0600 averaged 108 per cent. This is not a design possibility, and thus,

the assumption that resource usage by a job is uniform is not entirely accurate. Although an occasional anomaly exists, the uniform distribution of the computer resources usage by the subroutine UNIFORM, as discussed in the STOTAL section, proved to be more representative of the actual workload than when UNIFORM was not used.

Overhead Jobs. The jobs categorized as overhead consumed a significant portion of the computer system resources. For the days investigated, an average of 11.1 per cent of the total CRUs were produced by the overhead jobs. The actual average may even be higher since low averages on several days apparently do not reflect the full use of the JANUS program, which is a major consumer of computer resources. There are several reasons for the JANUS data to be missing; the data may have been recorded on the Dayfile outside of the 0800-1600 investigation or the accounting system may have failed to report the resource usage.

Values of the overhead CRUs ranged from a low of 0.1 per cent on 26 October to a high of 41.5 per cent on 25 October for the 0800 to 1600 investigation. A study of the investigation days for a 24 hour period revealed that an average of 19.9 per cent of the CRUs were produced by the overhead jobs. The values ranged from 2.1 per cent on 21 October to 36.9 per cent on 15 September.

Largest Jobs. A relatively small number of jobs consumed a major portion of computer system resources. Ten per cent of the jobs, which included overhead jobs, used an average of 69.0 per cent of all the CRUs produced during the days investigated. The range was from a low of 63.9 per cent on 13 October to a high of 76.6 per cent on 25 October.

Organization Utilization. The utilization of the computer resources by organization was measured in terms of CRUs produced. The results were relatively similar to the authorized per cent of utilization with two exceptions. The Computer Center used a much larger share of computer resources, the reason for this indication, as discussed earlier, is that jobs submitted at the Computer Center are not distinguishable by organization. All the overhead jobs and those jobs not specifically identified as an organization job are grouped into the Computer Center category. The AFIT organization also exceeded authorized usage, although an investigation indicated that the time period studied involved a peak period of activity for the academic institution.

Table IX presents the overall results of the 17 days which were investigated.

CPU Utilization. The first performance measure to be discussed is CPU utilization. A representative sampling covering several days of CPU utilization is presented in Table X. As may be observed in the table, there is a wide range in the CPU utilization.

During an eight hour period, the CPU utilization averaged 71.1% per hour, and varied from a high utilization of 83.5% per hour to a low utilization of 61.7% per hour. Two parameters were observed to be proportional to CPU utilization. The central processor execution time (CPTIME) for jobs with high CPU utilization was, not unpredictably, as much as one-fourth higher than for jobs with low CPU utilization. It was also observed that the amount of PPU utilization appeared to be proportional to CPU utilization (Fig. 14).

TABLE IX

Average Per Cent Utilization

For 17 Investigation Days

	Intercom CRUS	Batch CRUS	Overhead CRUS	Total CRUS	Authorized CRUS
AFIT	9.7	0.0	10.2	19.8	12.0
ASD	0.1	0.0	1.0	1.0	2.5
AFFDL	15.3	0.0	9.9	25.2	49.0
AFHRL	0.1	0.0	2.7	2.9	3.0
AMRL	3.4	0.0	1.7	5.1	9.0
AFML	0.0	0.0	0.0	0.0	1.0
AFAPL	0.0	0.0	0.0	0.0	1.0
AFAL	3.6	0.0	2.2	5.8	21.0
AFWAL	0.1	0.0	0.0	0.1	1.0
Computer Center and all others	5.3	2.2	32.7	40.2	0.5
Totals	37.9	2.2	60.9	100.0	100.0

TABLE X

CPU UTILIZATION SUMMARY

WORKLOAD PARAMETERS (AVERAGE VALUES)	11 OCT	30 SEP	26 OCT	27 OCT	13 OCT	15 SEP
CPTIME	21.67	29.15	22.89	23.26	34.10	32.75
PPTIME	35.36	83.76	33.71	68.21	73.83	67.96
TIMEIO	17.98	1056.61	18.65	23.53	17.13	1051.61
TOTCOST	26.75	36.52	28.02	33.40	35.81	33.64
KWS	767.43	1056.70	806.09	1022.81	1187.30	1052.19
CPOT	903.66	1103.85	933.11	1234.32	1176.29	1208.90
CMLOC	142858	180053	177853	76347	129083	126761
CPLOC	88.20	91.43	97.81	95.27	91.95	85.01
ROLLOUT	21.00	42.15	18.10	16.93	41.82	22.36
LATIME	47.28*	101.36	54.34	42.23*	41.49*	48.73*
INQTIME	598.58*	1533.99	525.54	645.79*	345.60*	514.96*
TAPEREQ	.060	.097	.031	.050	.074	.063
TAPEUSED	.053	.094	.031	.050	.071	.063

PERFORMANCE MEASURES (AVERAGE VALUES PER HR UNLESS STATED OTHERWISE)	11 OCT	30 SEP	26 OCT	27 OCT	13 OCT	15 SEP
JOBS/8 HRS	893	639	904	863	800	742
TURNAROUND/JOB	1400.71*	3176.84	1355.74	1567.16*	1224.57*	1551.52*
THRUPUT	111.63	79.88	113.00	107.88	100.00	92.75
BATCH JOBS	66.38	43.00	67.50	68.00	73.08	56.25
INTERCOM JOBS	40.63	36.00	44.63	38.50	34.58	35.75
CPU UTIL	61.70	62.00	70.00	72.90	83.50	82.70
PPU UTIL	110.50	119.00	102.60	191.30	187.00	171.60
CRUS	2771.70	2825.90	3073.30	3345.10	3150.90	3053.00
BATCH CRUS	1671.30	1362.10	1804.00	2060.60	2486.10	2155.90
INTERCOM CRUS	1100.40	1463.80	1269.30	1284.50	664.80	897.60
FIGURE OF MERIT	.000816*	.000087	.000587	.000460*	.003038*	.001340*

*Value based upon reduced data base for this parameter.

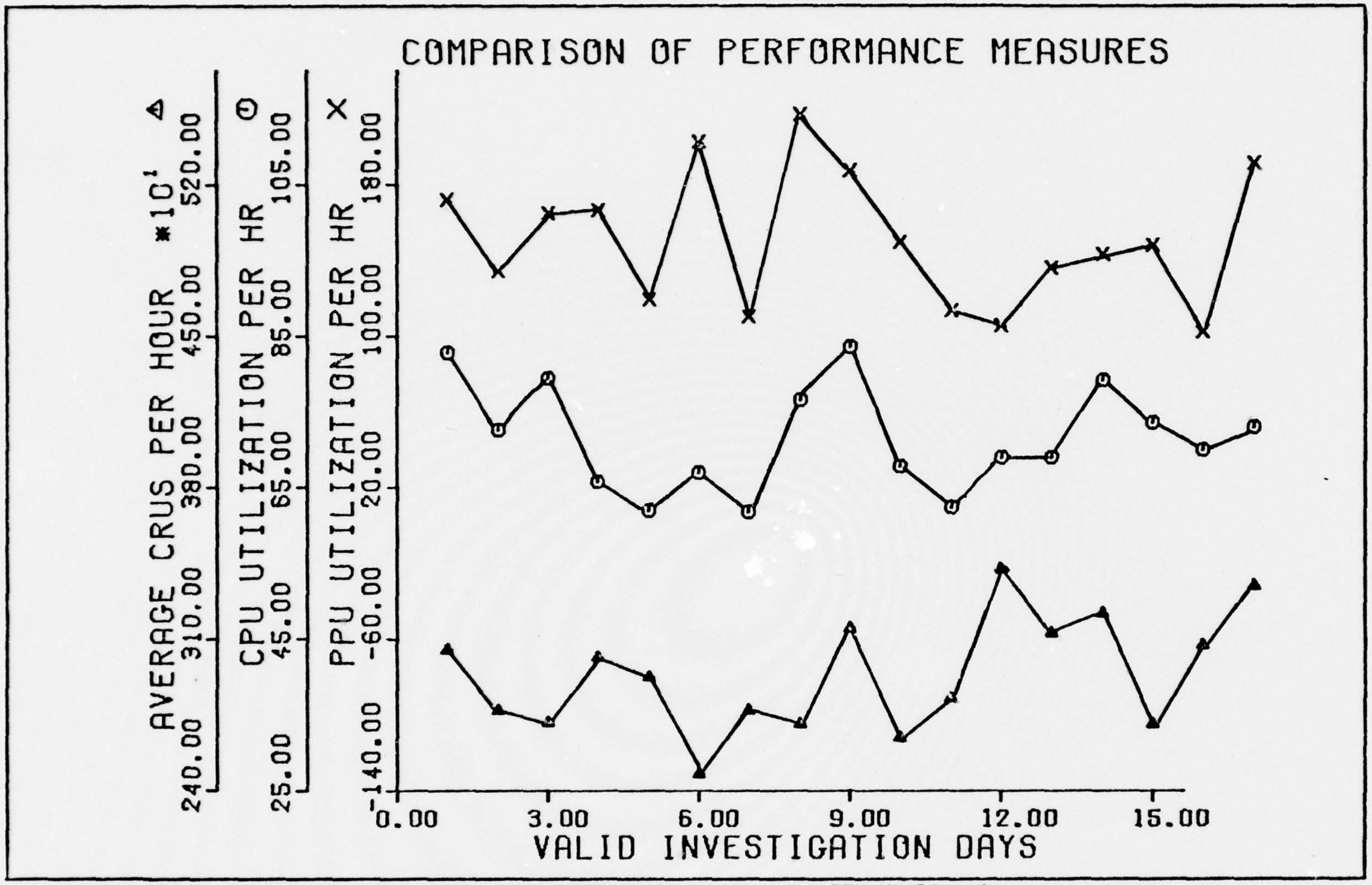


Fig. 14. Comparison of Average CRUs, CPU Utilization, and PPU Utilization

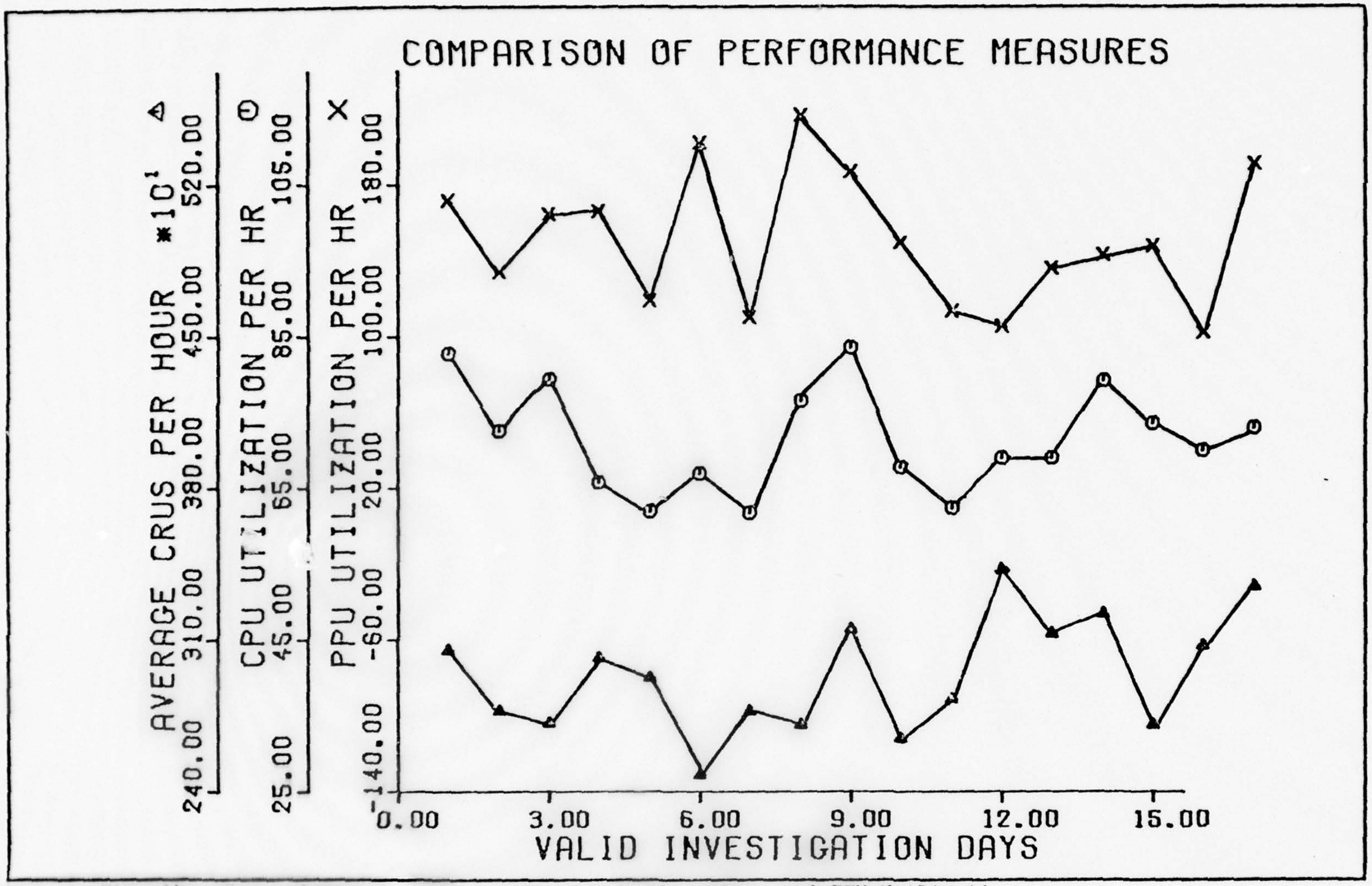


Fig. 14. Comparison of Average CRUs, CPU Utilization, and PPU Utilization

On high CPU utilization days, the following observations were made:

(1) The CRUs generated per hour are not noticeably larger than on medium utilization days (Fig. 15, Table X).

(2) The Figure of Merit is higher on those days when there is high CPU utilization (Fig. 15, Table X).

(3) The intercom throughput per hour is lower on high CPU utilization days (Fig. 16).

For comparison purposes, histograms depicting the highs and lows in CPU utilization are shown in Figures 17 and 18. Each histogram portrays the utilization for a 24 hour period; the 0800 to 1600 time interval is designated from 9 to 16 on the figures. The 30th of September represents a relatively low computer performance, while 13 October represents the higher performance. In Appendices E and H, the workload is characterized for 30 September and 13 October in terms of the probability density function and statistical analysis of each workload parameter.

Throughput. The next performance measure to be discussed is throughput. In Table XI, the range of throughput for the days investigated is presented. During the time-period 0800 to 1600, the number of jobs processed per hour averages 103.15, with a range from 126 to 65 jobs per hour. It had been anticipated that the larger jobs, which consume more resources, would tend to keep throughput time lower than would the smaller jobs. A comparison of these types of jobs indicated that on slower throughput days, such parameters as CPTIME, PPTIME, TIMEIO, TOTCOST, and KWS were generally larger, and thus, the result was a generally slower turnaround time and slower throughput per job.

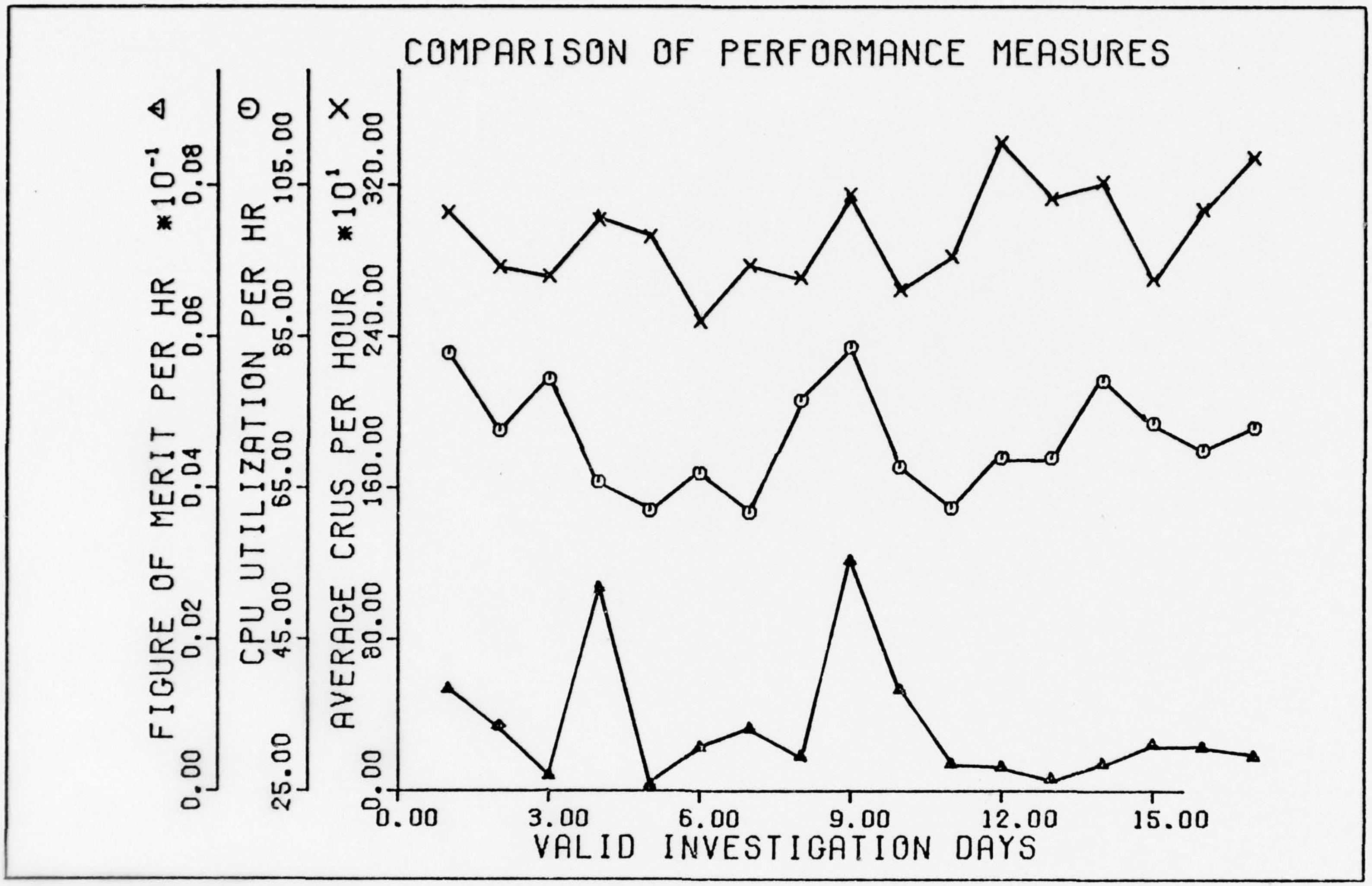


Fig. 15. Comparison of Figure of Merit, CPU Utilization, and Average CRUs

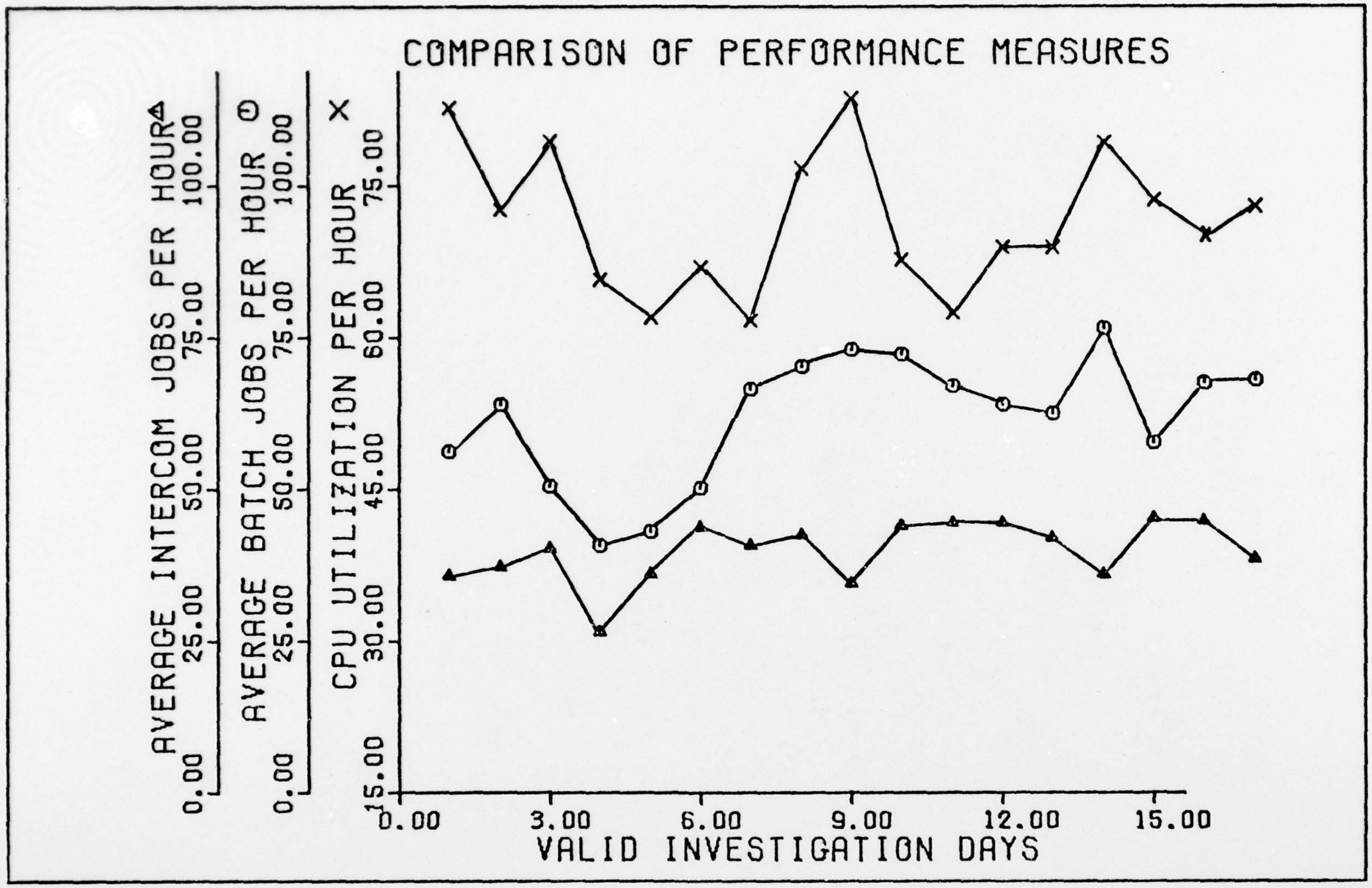


Fig. 16. Comparison of INTERCOM, Batch, and CPU Utilization

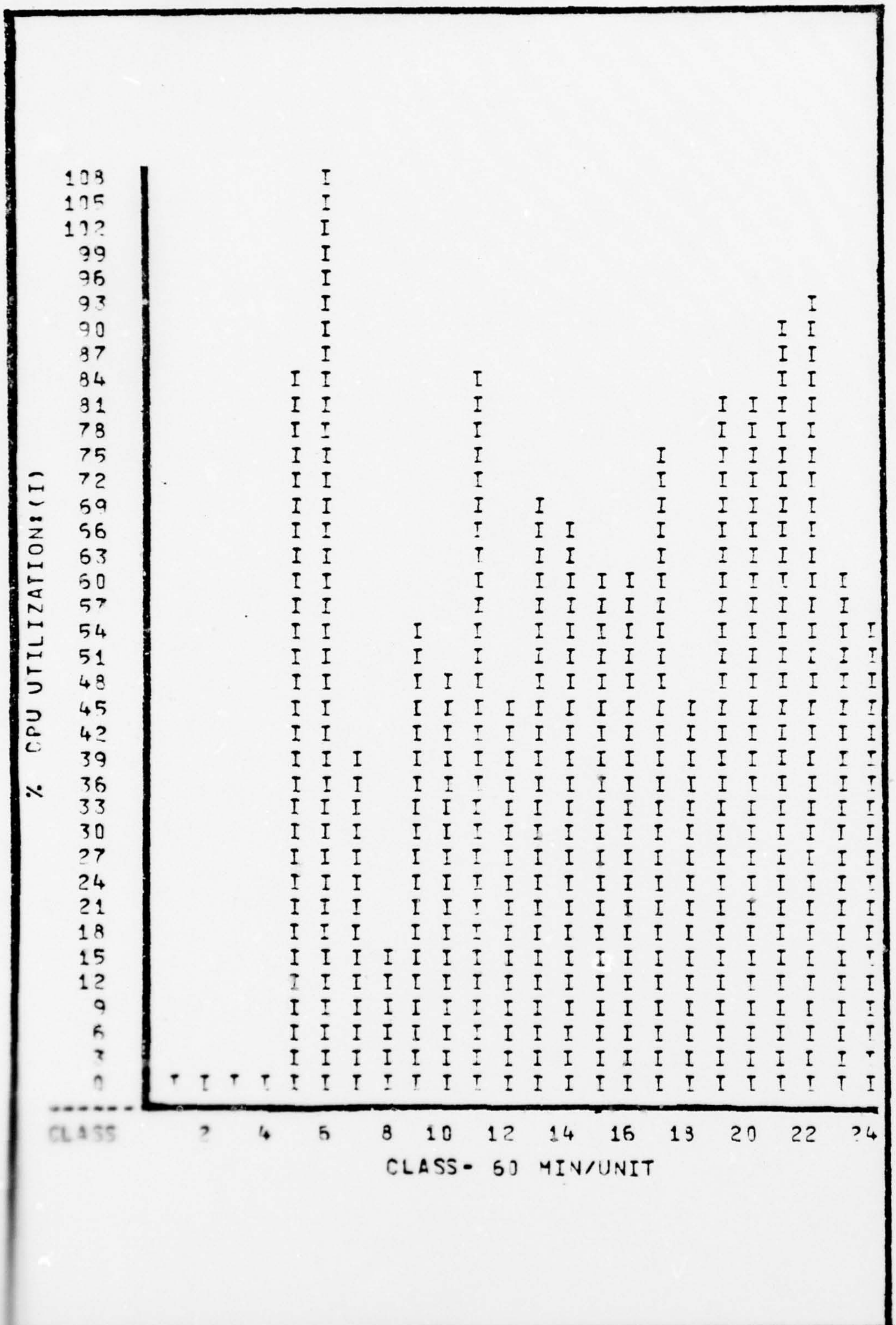


Fig. [illegible] CPU Utilization per Hour for 24 Hours (30 September)

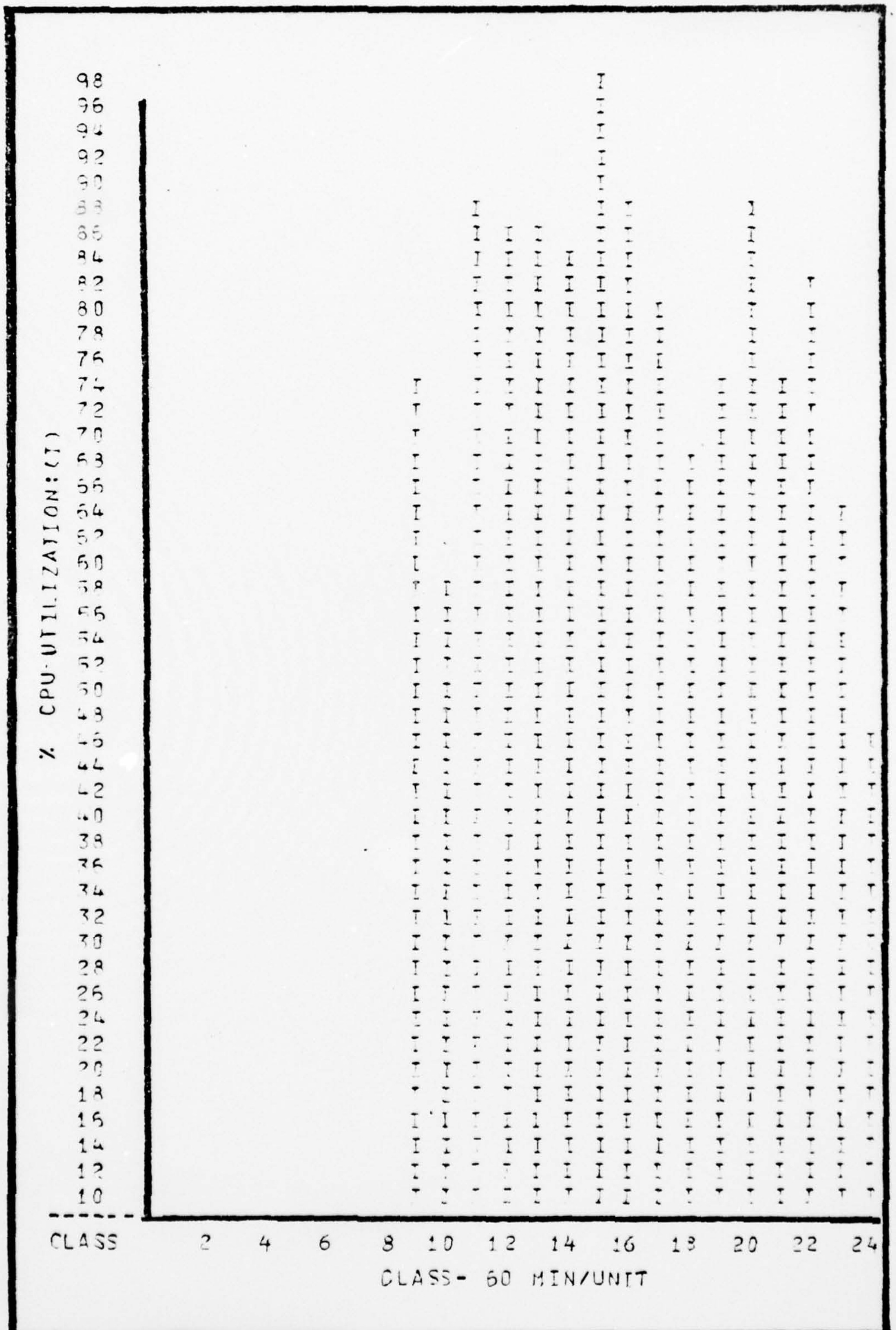


Fig. 18. CPU Utilization per Hour for 24 Hours (13 October)

TABLE XI

THROUGHPUT SUMMARY

WORKLOAD PARAMETERS (AVERAGE VALUES)	30 SEP	26 SEP	20 OCT	27 SEP	25 OCT	21 OCT
CPTIME	29.15	34.43	22.13	35.53	24.28	29.68
PPTIME	63.76	66.50	48.45	56.62	78.36	72.02
TIMEIO	1056.61	839.96	24.023	808.17	13.95	16.34
TOTCOST	36.52	32.49	31.00	27.66	24.085	29.442
KWS	1056.70	839.98	834.04	816.16	648.56	825.16
CPOT	1103.85	1142.60	1080.86	760.25	1045.16	964.20
CMLCC	180052	51228	187128	166231	191314	146109
CPLCC	91.83	96.86	95.44	67.57	90.32	96.42
ROLLOUT	42.15	41.68	22.33	16.19	22.13	16.17
LATIME	101.36	69.63	46.04	44.88*	40.06*	50.23
INQTIME	1533.99	1218.00	1651.24	322.38*	129.74*	831.63
TAPEREQ	.097	.093	.057	.073	.052	.056
TAPEUSED	.094	.092	.053	.067	.052	.053

PERFORMANCE MEASURES (AVERAGE VALUES PER HR UNLESS STATED OTHERWISE)	30 SEP	26 SEP	20 OCT	27 SEP	25 OCT	21 OCT
JOBS/8 HRS	639	728	825	834	1006	942
TURNAROUND/JOB	3176.84	2726.35	3081.54	790.37*	769.89*	1752.68
THRUPUT	79.88	91.00	103.13	104.25	125.75	117.75
BATCH JOBS	43.00	50.38	62.38	63.88	57.75	76.38
INTERCOM JOBS	36.00	40.25	42.00	37.13	45.25	39.38
CPU UTIL	62.00	79.30	68.90	72.60	73.50	79.20
PPU UTIL	182.20	164.80	136.30	134.70	148.10	143.50
CRUS	2825.90	2718.30	3126.80	2768.50	2701.30	3216.70
BATCH CRUS	1362.10	1633.70	1529.00	1849.40	1631.60	2177.70
INTERCOM CRUS	1463.80	1089.60	1597.80	919.10	1069.70	1039.0
FIGURE OF MERIT	.000087	.000199	.000168	.000876*	.000627*	.000335

*Value based upon reduced data base for this parameter.

The present policy of the Computer Center is to limit the size of jobs submitted during the day; as a result, the throughput is faster than if size of job were unrestricted. Further increases in throughput may be achieved if the restriction on size were strictly enforced or if the size limit were reduced further.

Throughput did not appear to be related to CPU utilization, although PPU utilization was higher on slow throughput days (Figs. 16, 19).

Figures 20 and 21 are histograms which depict the throughput for two of the days investigated, 30 September and 25 October. These two days represent respectively the slow and fast extremes in throughput. The workload for these days is characterized in Appendices E and D.

The total jobs, representing total batch and interactive jobs, was significantly lower on the 30th of September. It is interesting to note a common pattern of usage by the interactive jobs for all days investigated. The pattern showed that there was a surge of jobs from 0800-0900 (users arrival at work), from 1200-1300 (after lunch), and from 1500-1600 (large jobs entering the system due to management policy) (Figs.20, 21).

Turnaround. The third performance measure is based upon the average turnaround per batch job for the eight hour period from 0800 to 1600. The range of values is shown in Table XII.

The average machine turnaround time (ITURN) ranged from 1189.74 seconds (19.8 minutes) to 3176.835 seconds (53.0 minutes) for jobs which were basically the same size in terms of CPTIME, PPTIME, TIMEIO, and KWS. For all of the investigation days, the average turnaround time for a batch job was 28.3 minutes.

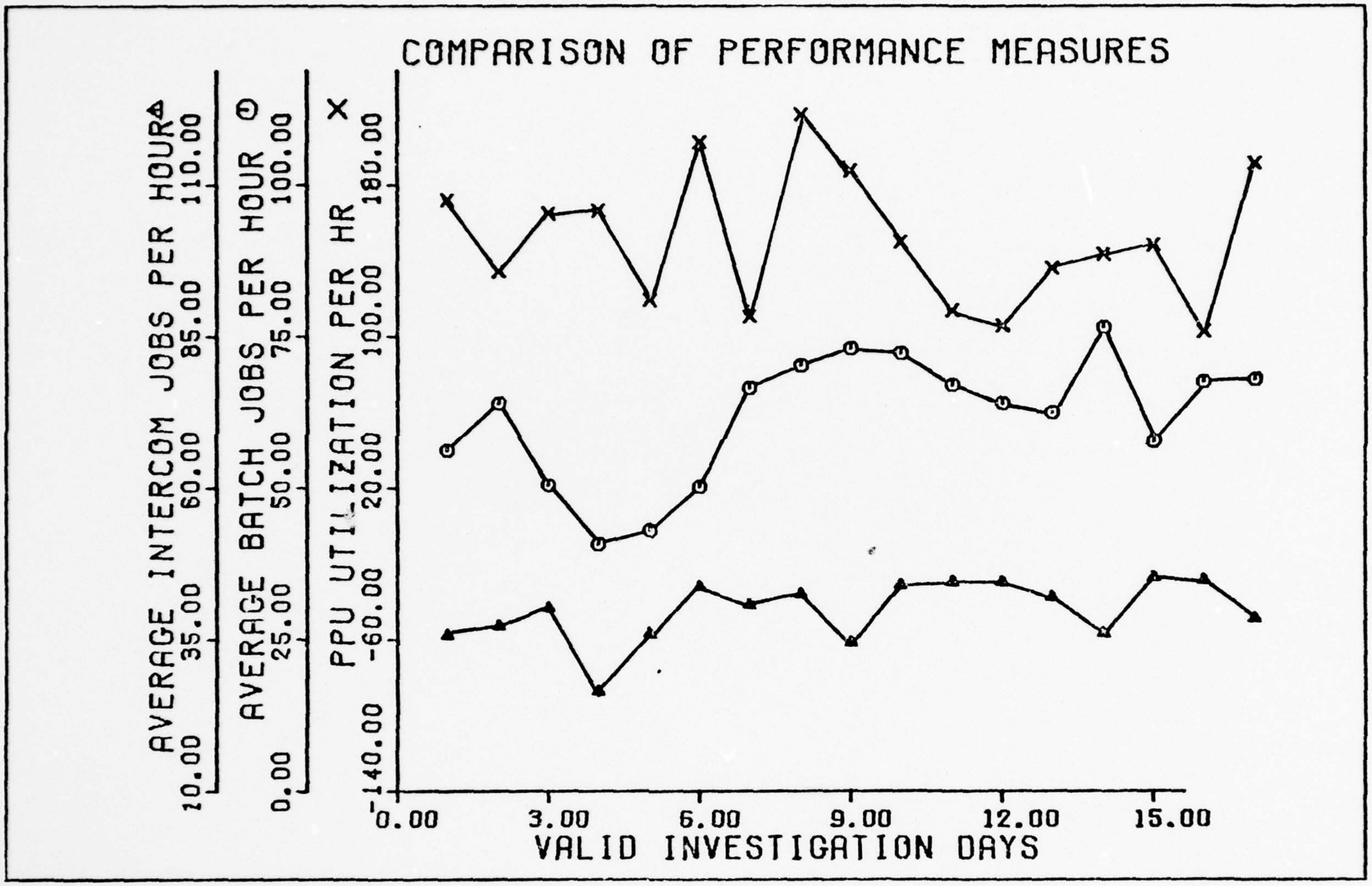


Fig. 19. Comparison of INTERCOM, Batch, and PPU Utilization

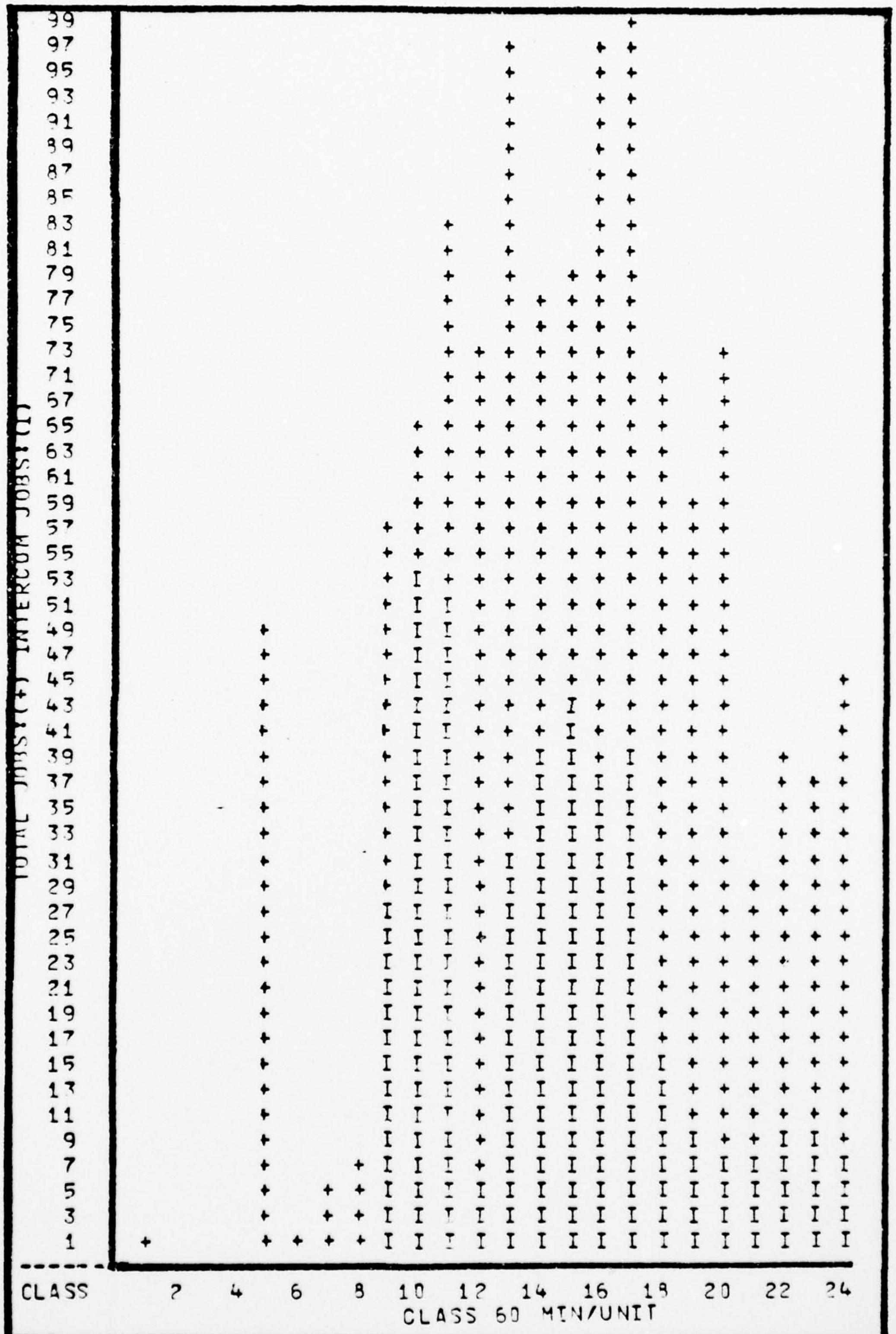


Fig. 20. Throughput per Hour for 24 Hours (30 September)

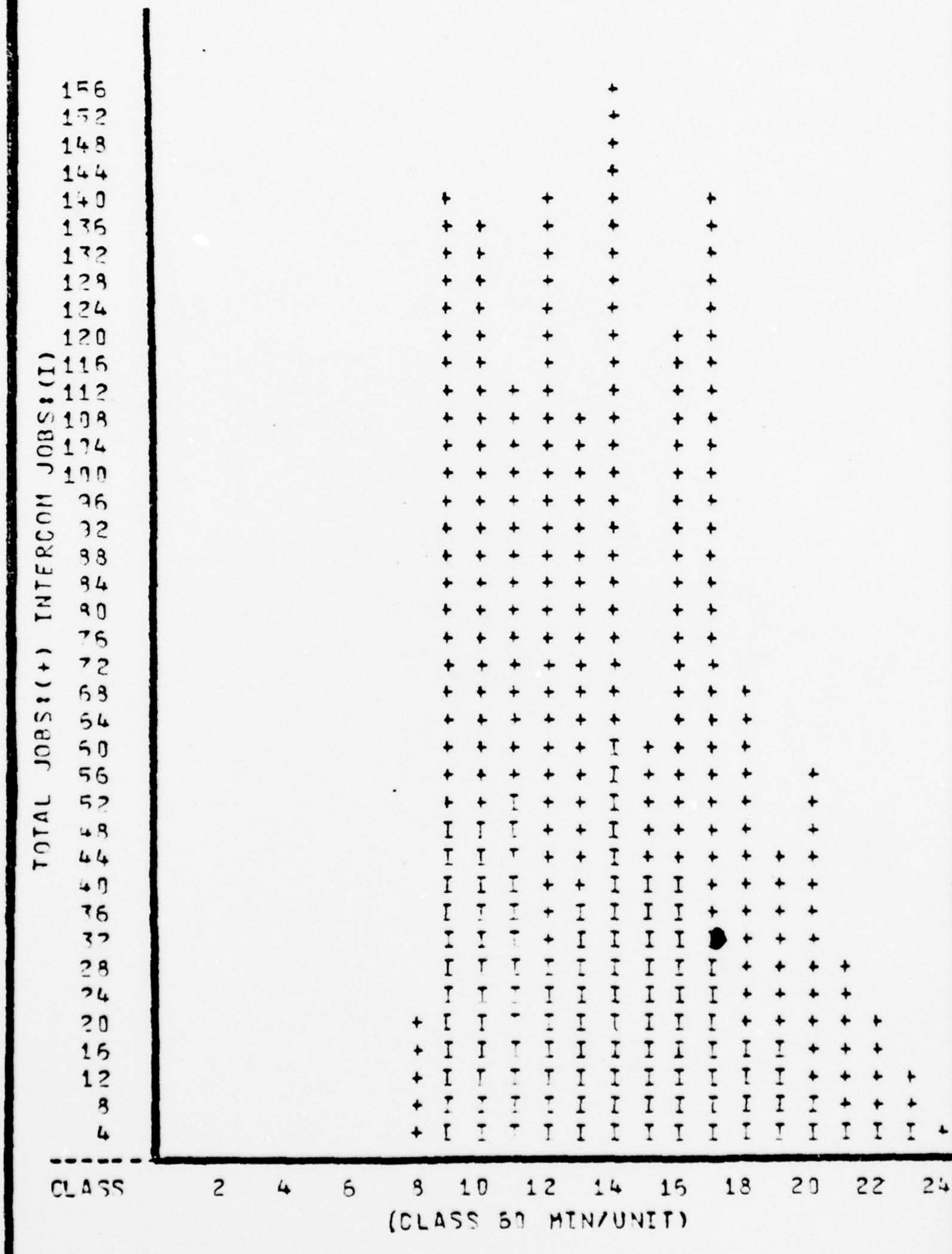


Fig. 21. Throughput per Hour for 24 Hours (25 October)

TABLE XII

TURNAROUND SUMMARY

WORKLOAD PARAMETERS (AVERAGE VALUES)	30 SEP	20 OCT	28 SEP	19 OCT	21 OCT	14 OCT
CPTIME	29.15	22.13	34.43	23.12	29.60	22.60
PPTIME	83.78	48.45	66.50	34.53	72.02	62.36
TIMEIO	1056.61	24.02	839.96	22.95	16.34	15.49
TOTCOST	36.52	31.00	32.49	31.43	29.44	24.96
KWS	1056.70	834.04	839.98	888.88	826.16	701.93
CPOT	1103.85	1080.86	1142.59	863.50	964.20	777.24
CMLOC	180053	187128	51228	129792	146110	124030
CPLOC	91.83	95.44	94.86	97.25	96.42	94.85
ROLLOUT	42.15	22.33	41.63	27.50	16.17	17.37
IATIME	101.36	46.04	69.83	49.52	50.23	44.50
INQTIME	1533.99	1651.24	1218.00	1310.99	931.63	576.55
TAPEREQ	.097	.057	.093	.048	.056	.062
TAPEUSED	.094	.053	.092	.047	.053	.058
PERFORMANCE MEASURES (AVERAGE VALUES PER HR UNLESS STATED OTHERWISE)						
JOBS/8 HRS	639	825	728	899	942	936
TURNAROUND/JOB	3176.84	3081.54	2726.35	2544.44	1752.68	1189.73
THRUPUT	79. 8	103.13	91.00	112.38	117.75	117.00
BATCH JOBS	43.00	62.38	50.38	63.88	76.38	72.25
INTERCOM JOBS	36.00	42.00	40.25	44.50	39.88	44.00
CPU UTIL	62.00	61.60	79.30	68.90	79.20	67.70
PPU UTIL	182.20	136.30	164.80	105.40	143.50	149.80
CRUS	2825.90	3126.80	2718.30	3422.80	3216.70	2644.10
BATCH CRUS	1362.10	1529.00	1633.70	1913.60	2177.70	1776.80
INTERCOM CRUS	1463.80	1597.80	1084.60	1509.50	1039.00	867.30
FIGURE OF MERIT	.000087	.0000168	.000199	.000324	.000355	.001693

The slower turnaround could be a result of the computer system being saturated with jobs. As an example, on slower turnaround days, the time spent by a job in the input queue waiting for processing (INQTIME) and the number of control points filled for a job (CPLOC) were larger than for jobs on days with faster turnaround. Faster turnaround appeared to be proportional to faster throughput time and inversely proportional to CRUs per hour (Fig. 22). In addition the slower turnaround days also tended to have a lower Figure of Merit; this indicates lower priority of service for small jobs on slow turnaround days (Fig. 23). Future investigations into intercom turnaround time and response time may further explain these relationships.

In Figures 24 and 25, the extremes of average turnaround per day are presented for comparison. Lowest average turnaround was on 30 September and highest average turnaround was on 14 October.

The backlog on the turnaround in both histograms reflects the number of jobs which enter the computer system at the 0800, 1200, and 1600 periods.

In Appendices E and F, the workload is characterized statistically for the two days representing the class of turnaround.

CRUs. The class of workload to be discussed in this section represents the CRU performance measure. As may be seen in Table XIII, there is a wide range of total CRUs for the eight hour period on the various days investigated.

For the overall investigations, average CRUs per hour for the CYBER 74 computer system was 3001. 21; the range of CRUs was from 2477.93 to 3422.78. In terms of this class of workload, the jobs during

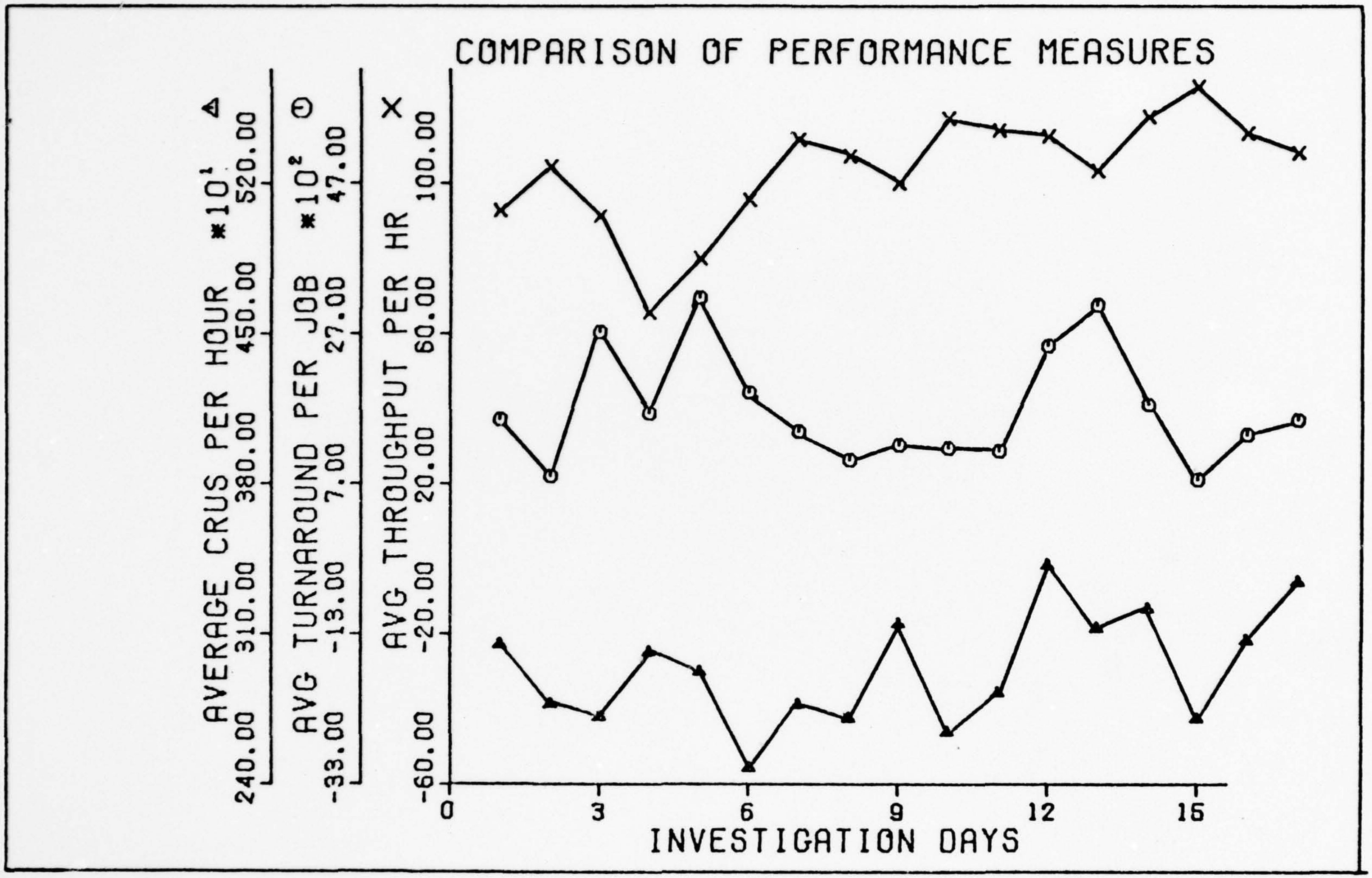


Fig. 22. Comparison of CRUs, Turnaround, and Throughput

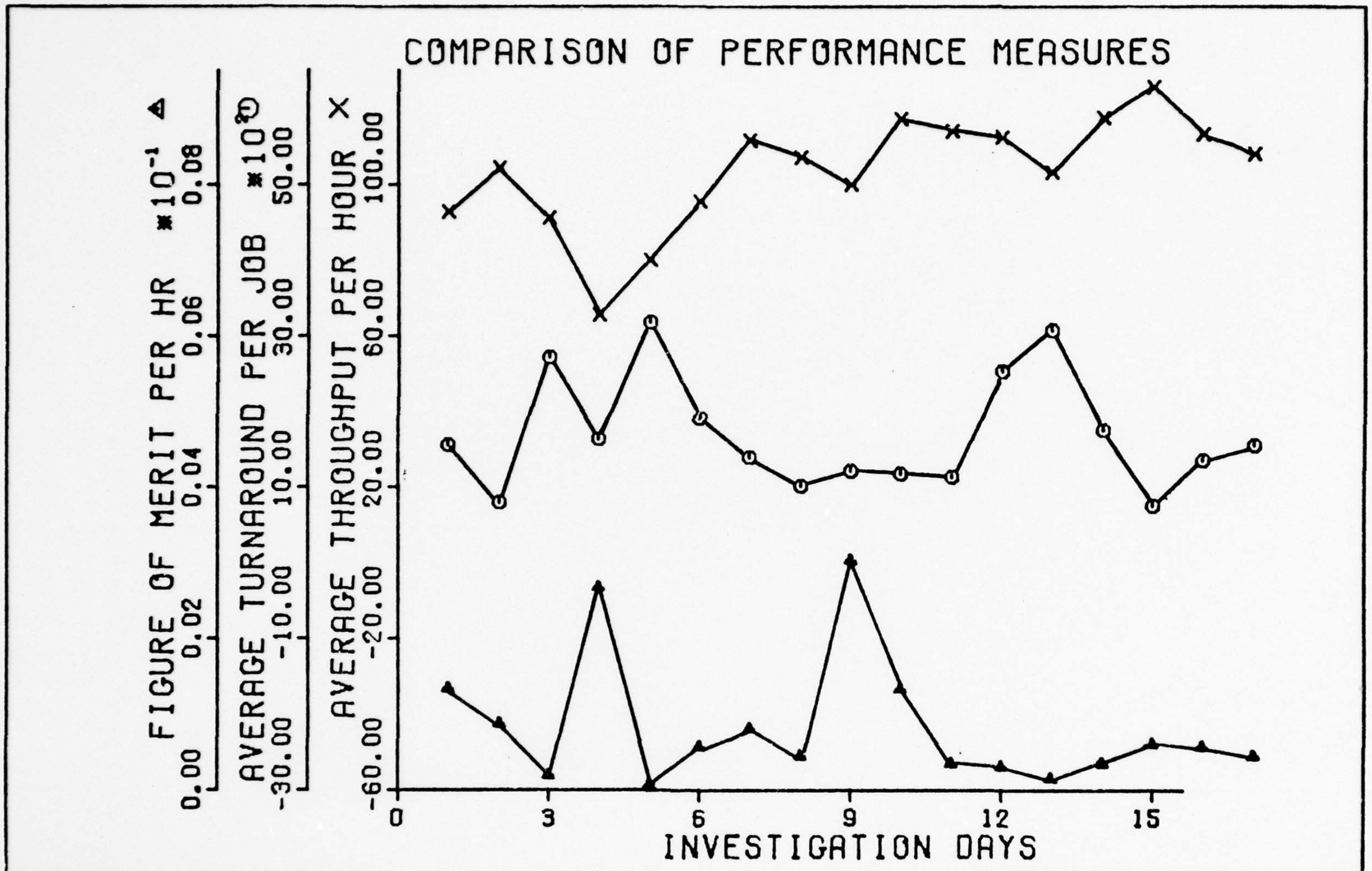


Fig. 23. Comparison of Figure of Merit, Turnaround, and Throughput

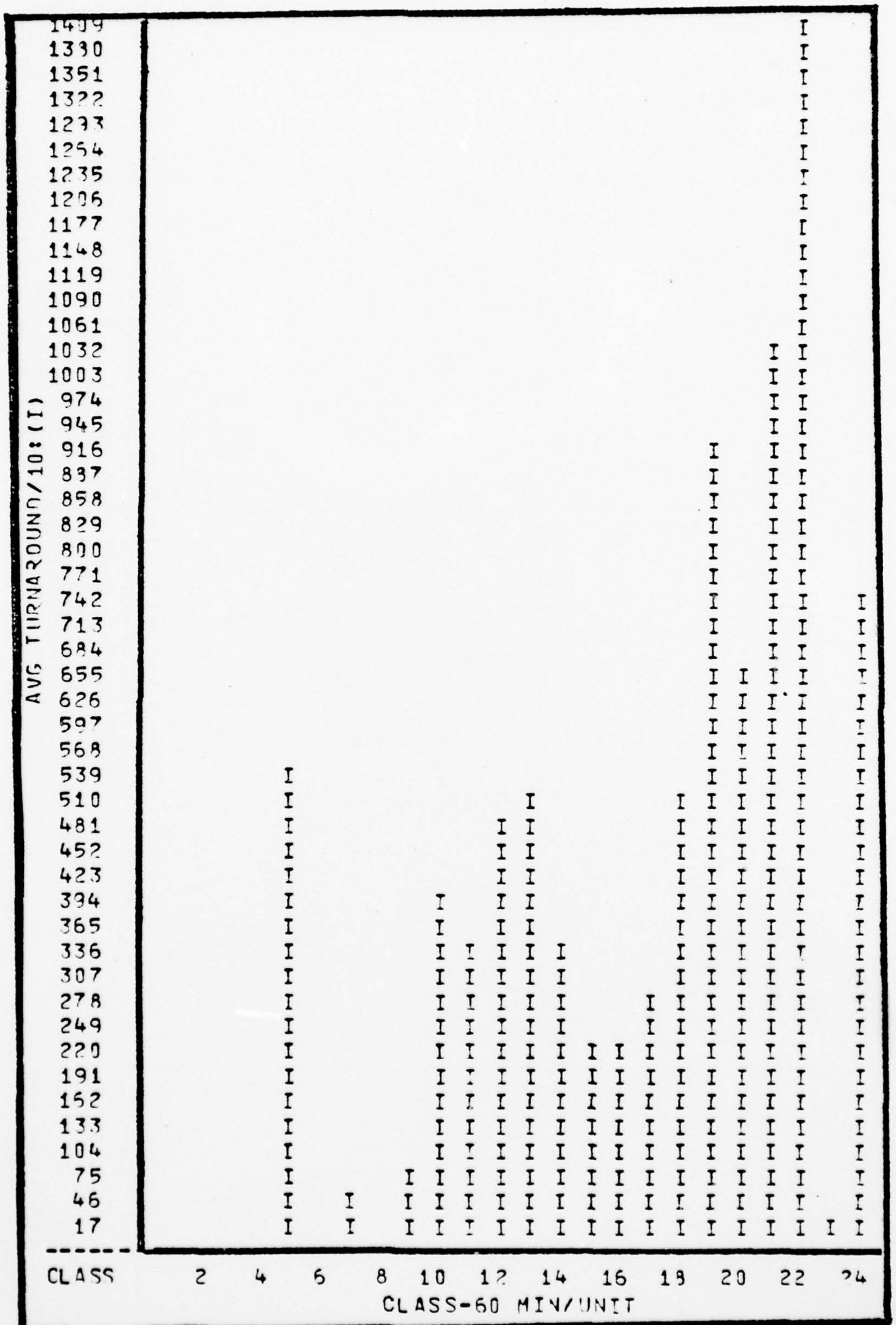


Fig. 24. Average Turnaround per Hour for 24 Hours (30 September)

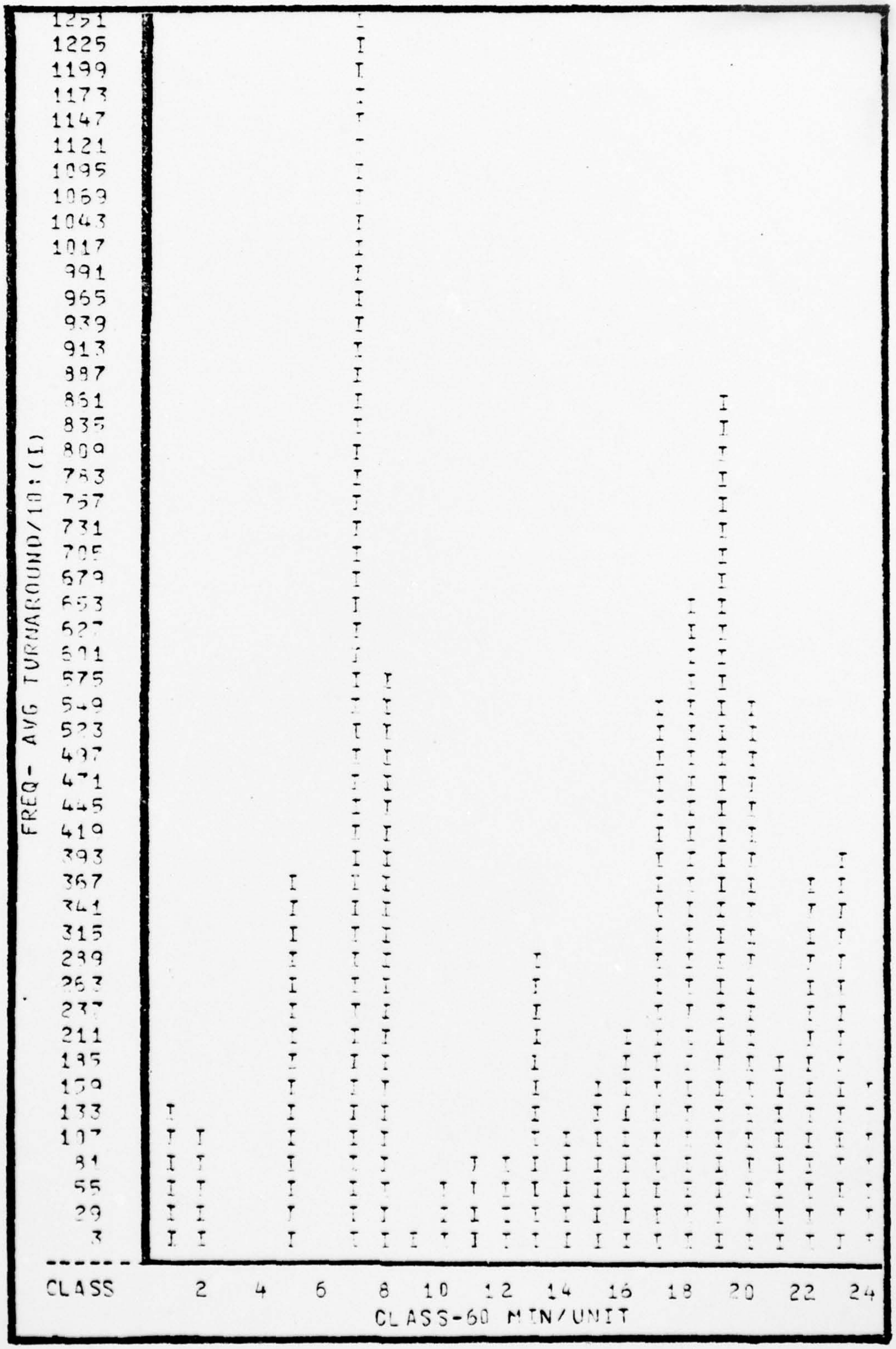


Fig. 25. Average Turnaround per Hour for 24 Hours (14 October)

TABLE XIII

CRU SUMMARY

WORKLOAD PARAMETERS (AVERAGE VALUES)	7 OCT	14 OCT	29 SEP	26 OCT	19 OCT	27 OCT
CPTIME	27.31	22.60	35.17	22.89	23.12	23.26
PPTIME	98.03	62.36	95.72	33.71	34.53	68.21
TIMEIO	881.05	15.49	1397.72	18.65	22.95	23.53
TOTCOST	31.46	29.96	43.55	28.02	31.43	33.40
KWS	881.25	701.93	1397.78	806.09	888.58	1022.81
CPCT	804.03	777.24	1158.486	933.11	863.50	1234.32
CMLOC	149777	124030	202640	174853	129792	76735
CPLOC	61.21	94.85	92.19	97.81	97.25	95.27
ROLLOUT	22.46	17.37	61.29	18.10	27.50	16.83
IATIME	65.35	44.50	49.92*	54.34	49.523	42.22*
INQTIME	808.29	576.55	799.85*	525.54	1310.99	645.79*
TAPEREQ	.059	.062	.078	.031	.048	.050
TAPEUSED	.055	.058	.076	.031	.047	.040

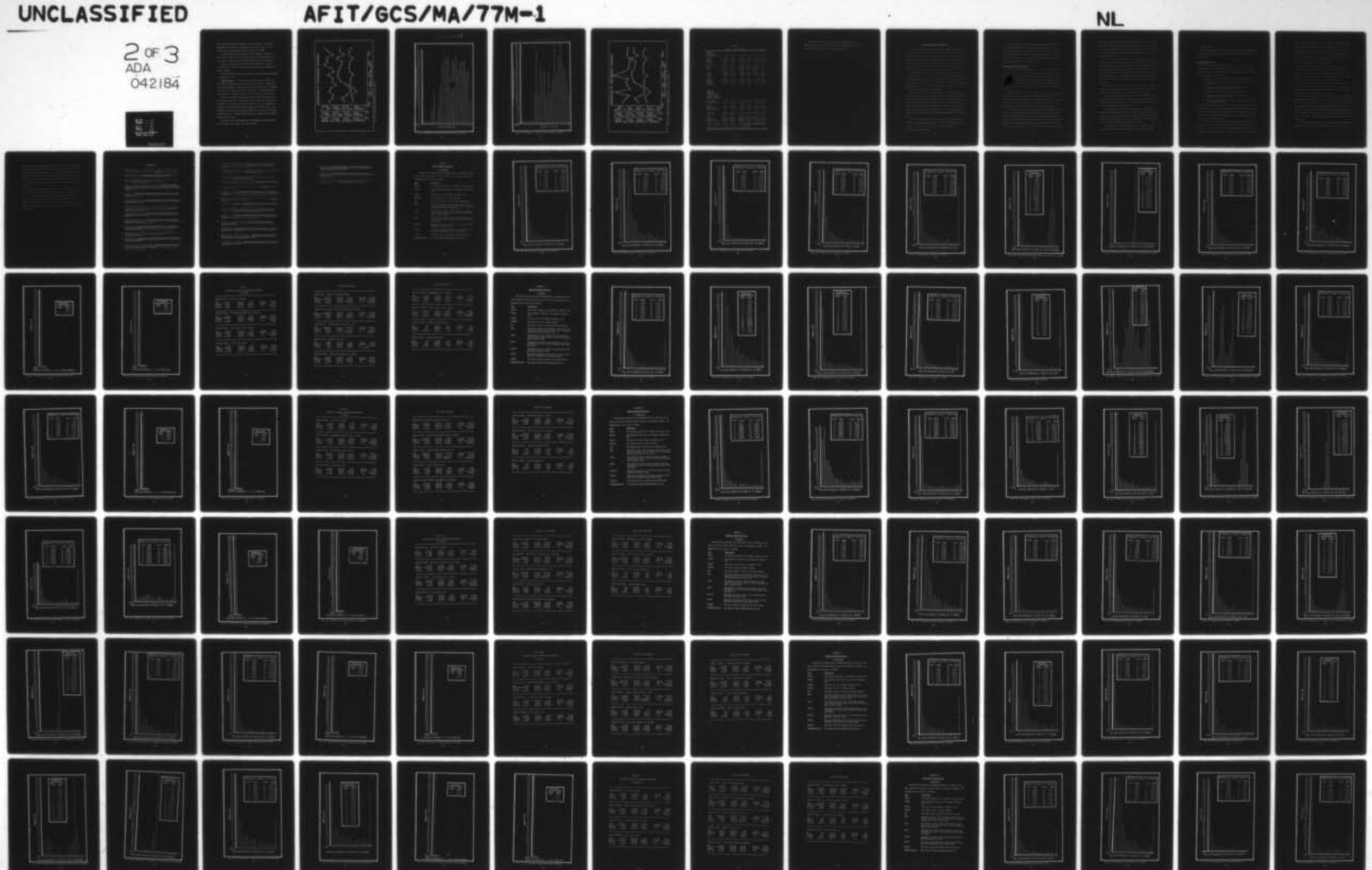
PERFORMANCE MEASURES (AVERAGE VALUES PER HR UNLESS STATED OTHERWISE)						
JOBS/8 HRS	764	936	525	904	899	863
TURNAROUND/JOB	1912.84	1189.73	1638.52*	1355.74	2544.41	1567.16*
THRUPUT	95.50	117.00	65.63	113.00	112.38	107.88
BATCH JOBS	50.13	72.25	40.75	67.50	63.88	68.00
INTERCOM CRUS	43.63	44.00	26.50	44.63	44.50	38.50
CPU UTIL	66.90	67.70	65.70	70.00	68.90	72.90
PPU UTIL	202.70	149.80	166.00	102.60	105.40	191.30
CRUS	2477.90	2644.10	3017.60	3073.30	3422.80	3345.10
BATCH CRUS	1680.00	1776.80	1849.80	1804.00	1913.60	2059.20
INTERCOM CRUS	797.90	867.30	1167.80	1269.30	1509.50	1285.90
FIGURE OF MERIT	.000589	.001693	.002678*	.000587	.000324	.000460*

*Value based upon reduced data base for this parameter.

AD-A042 184 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO SCH--ETC F/G 9/2
WORKLOAD CHARACTERIZATION AND MEASUREMENT OF THE CDC CYBER 74 C--ETC(U)
MAR 77 J R BEAR

UNCLASSIFIED AFIT/GCS/MA/77M-1 NL

2 OF 3
ADA
042184



the investigation period appeared to be quite similar. The workload parameter, control points, appeared to be related to CRUs, as more control points were assigned on those days with high CRUs.

As seen in Figure 26, the number of CRUs appears to depend upon the number of batch jobs and upon the number of CRUs generated by the batch jobs. Intercom jobs and intercom CRUs are relatively constant.

Figures 27 and 28 are histograms which represent the two days, 19 October and 7 October, on which the low and the high CRU generation values occurred.

The CRU performance class of workload is statistically characterized in Appendices G and B.

Figure of Merit. The final workload characterization class to be discussed is Hellerman's Figure of Merit for batch jobs. A representative sampling depicting the Figure of Merit for several days is presented in Table XIV. As discussed earlier, nine of the 17 investigation days were based upon a reduced data base due to missing input queue times.

There is a relatively narrow range in the Figure of Merit; all of the values are very low. The average Figure of Merit for the eight days was .000500. This would indicate that very little if any consideration is given to providing higher priority service to the short jobs. It was observed that high batch throughput appeared to be related to the Figure of Merit (Fig. 29). Larger Figures of Merit are related to faster turnaround time (Fig. 23).

Histograms are not presented for this performance measure because of the extremely low values which were obtained.

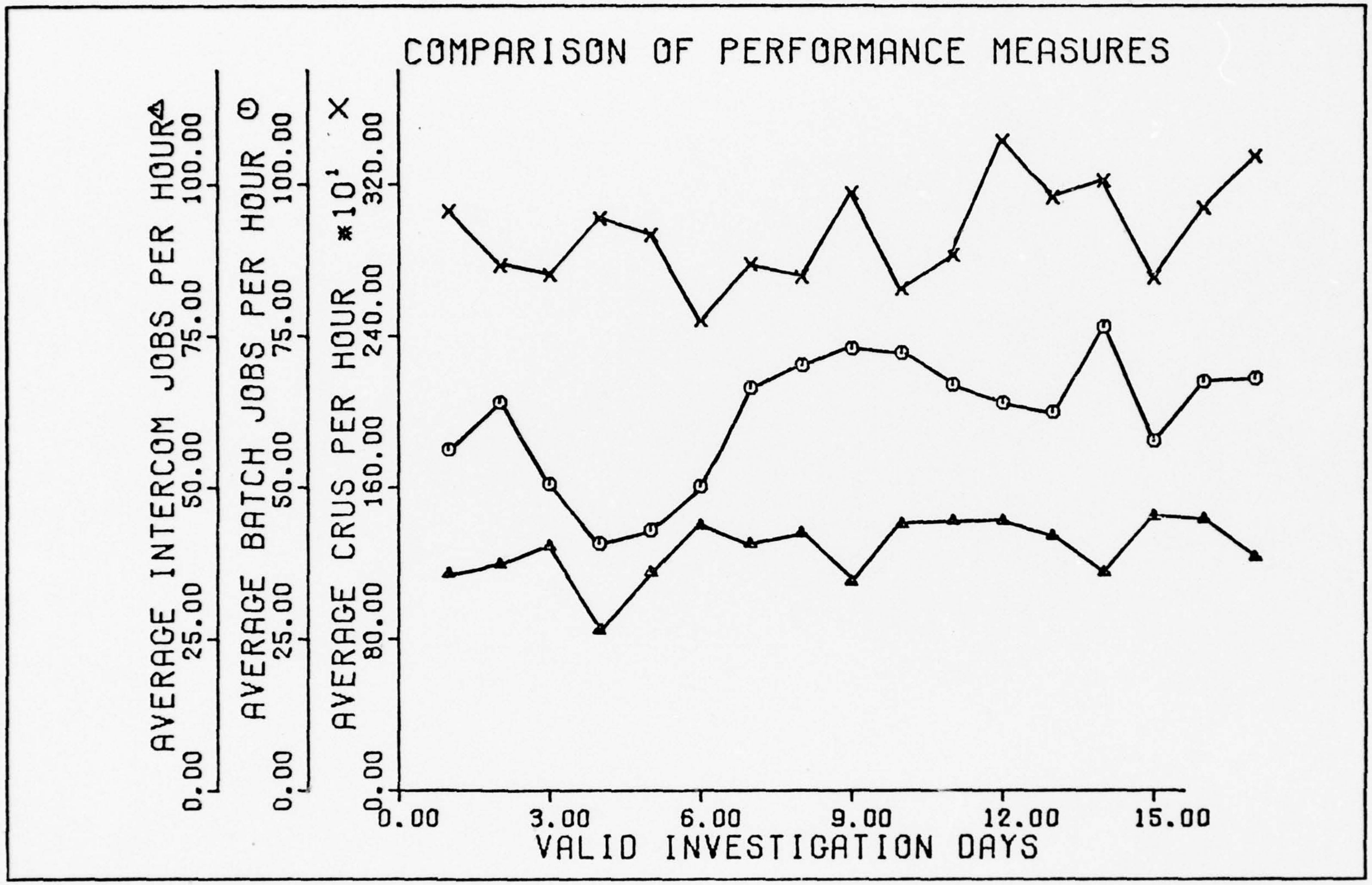


Fig. 26. Comparison of INTERCOM, Batch, and CRUs

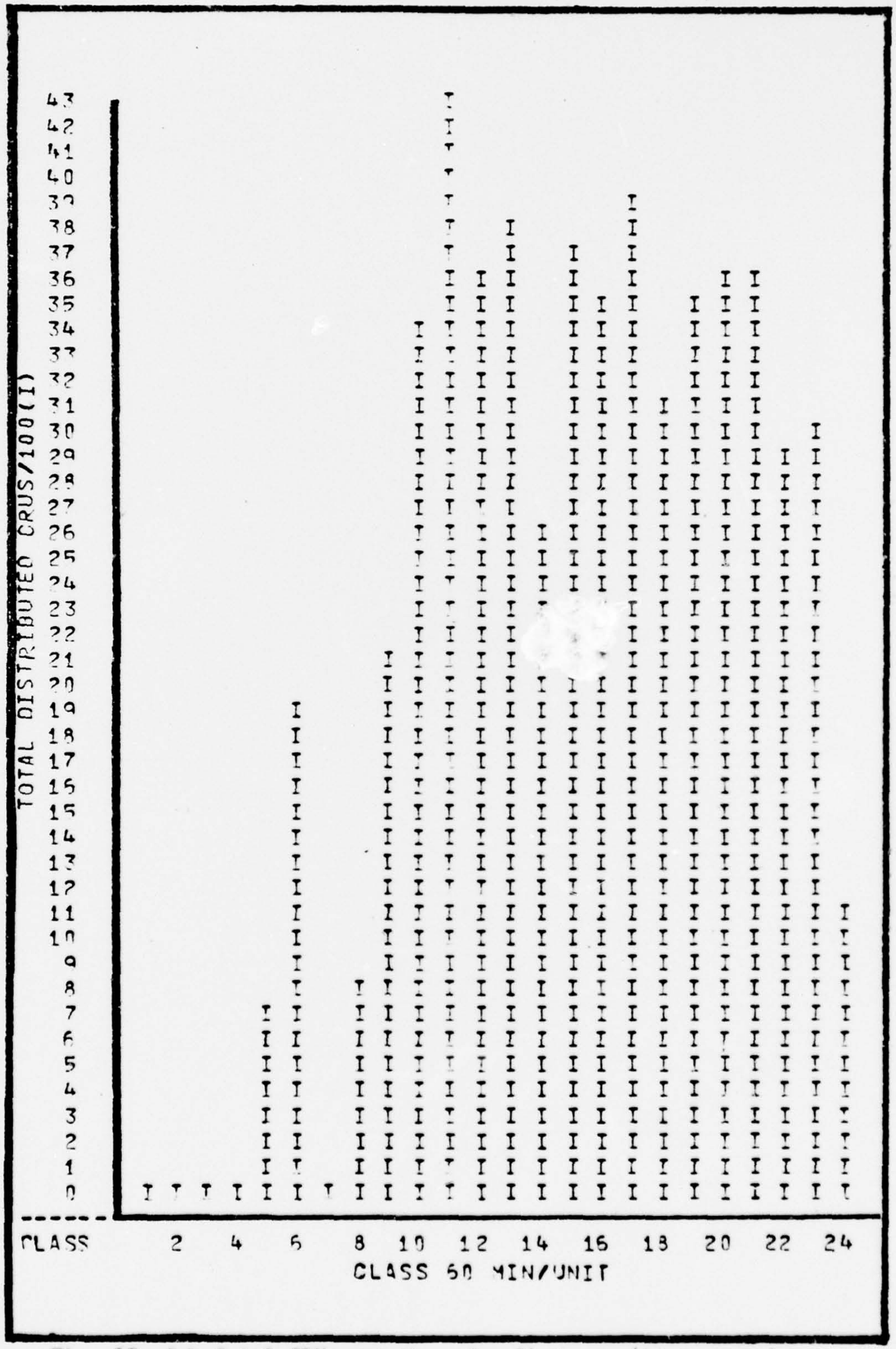


Fig. 27. Job Total CRUs per Hour for 24 Hours (19 October)

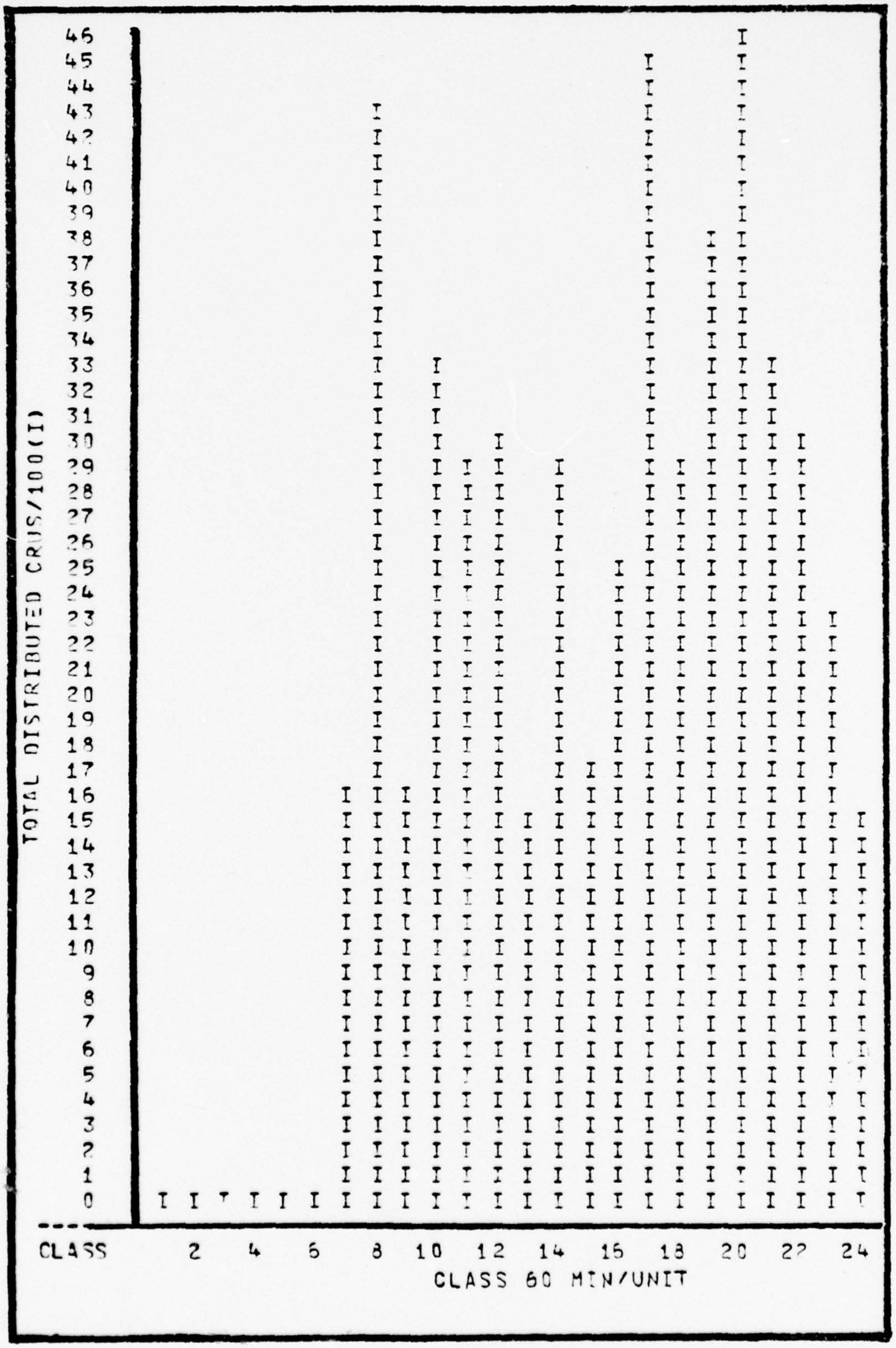


Fig. 28. Job Total CRUs per Hour for 24 Hours (7 October)

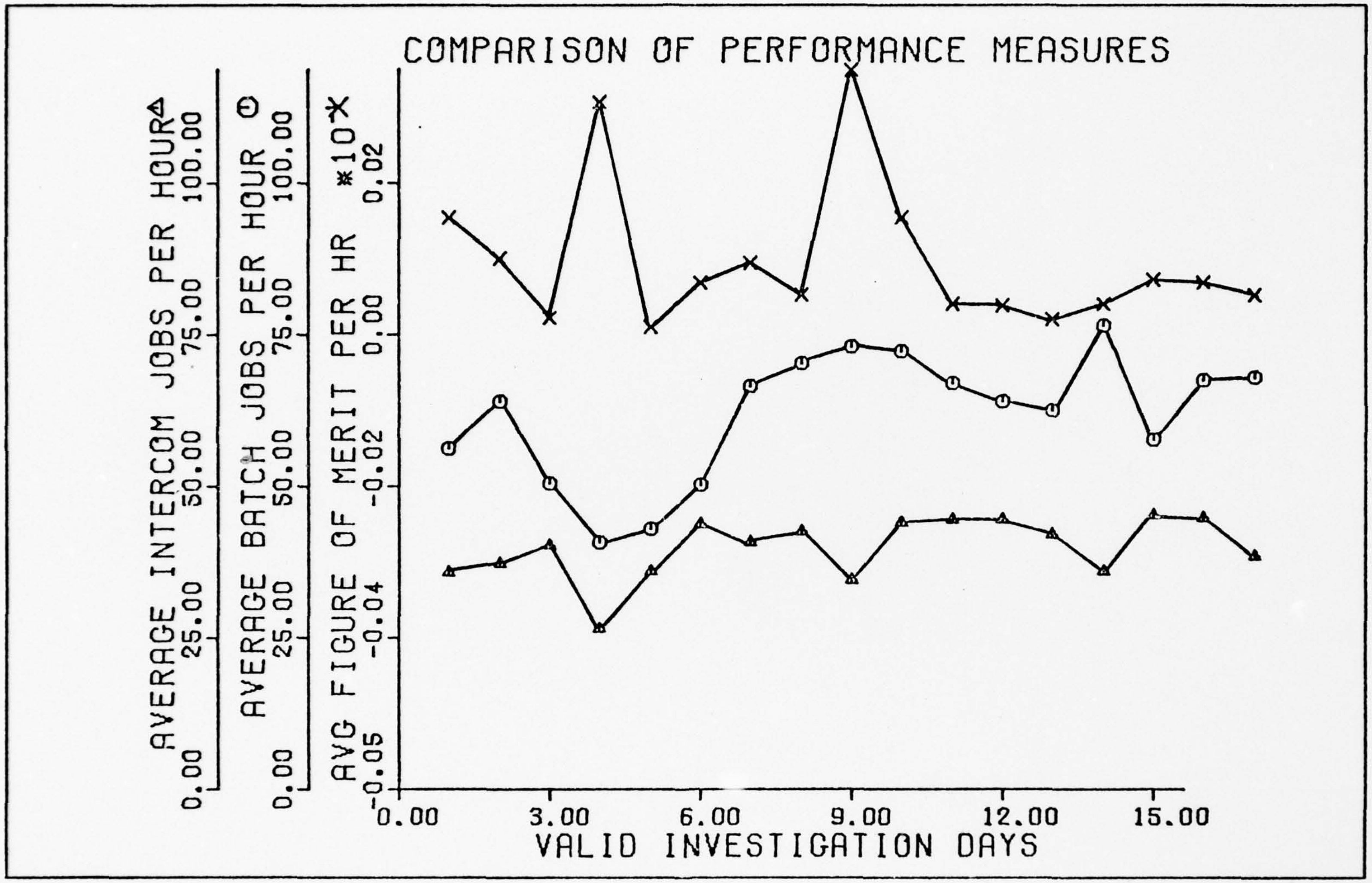


Fig. 29. Comparison of INTERCOM, Batch, and Figure of Merit

TABLE XIV

FIGURE OF MERIT SUMMARY

WORKLOAD PARAMETERS (AVERAGE VALUES)	30 SEP	20 OCT	19 ICT	7 OCT	13 OCT	14 OCT
CPTIME	29.15	22.13	23.12	27.31	34.10	22.60
PPTIME	83.76	48.45	34.53	98.03	73.83	62.36
TIMEIO	1056.61	24.02	22.95	881.05	17.13	15.49
TOTCOST	36.52	31.00	31.43	31.46	35.81	24.96
KWS	1056.70	834.04	888.88	881.25	1187.30	701.93
CPOT	1103.85	1080.86	863.50	804.03	1176.29	777.24
CMLOC	180053	187128	129792	149777	129083	124030
CPLOC	91.83	95.44	97.25	61.21	91.95	94.85
ROLLOUT	42.15	22.33	27.50	22.46	41. 2	19.37
LATIME	101.36	46.04	49.52	65.35	41.49*	44.50
INQTIME	1533.99	1651.24	1310.99	808.29	345.60*	576.55
TAPEREQ	.097	.057	.048	.059	.074	.062
TAPEUSED	.094	.053	.047	.085	.071	.058
PERFORMANCE MEASURES (AVERAGE VAOUES PER HR UNLESS STATED OTHERWISE)						
JOBS/8 HRS	639	825	899	764	800	936
TURNAROUND/JOB	3176.84	3081.54	2544.41	1912.85	1224.57*	1189.74
THRUPUT	79.88	103.13	112.38	95.50	100.00	117.00
BATCH JOBS	43.00	62.30	63.88	50.13	73.08	72.25
INTERCOM JOBS	36.00	42.00	44.50	43.63	34.58	44.00
CPU UTIL	62.00	61.60	68.90	66.90	83.50	67.70
PPU UTIL	182.20	136.30	105.40	202.70	187.00	149.80
CRUS	2825.90	3126.80	3422.78	2477.90	3150.90	2644.10
BATCH CRUS	1362.10	1529.00	1913.60	1680.00	2486.10	1776.80
INTERCOM CRUS	1463.80	1597.80	1509.50	797.90	664.80	867.30
FIGURE OF MERIT	.000087	.000168	.000324	.000589	.003038*	.001693

*Value based on reduced data base for this parameter.

In Appendices E and F, the workload is characterized for 30 September and 14 October in terms of the probability density function and a statistical analysis of each workload parameter.

V. Conclusions and Recommendations

The three objectives of this thesis were to analyze the accounting data, to characterize the workload, and to measure the impact of variations in the workload upon the performance of the CYBER 74 computer system. These objectives have been accomplished by use of the CLARA data. Some refinements of the accounting data capabilities would make possible a more comprehensive description of the CYBER 74, and a more accurate description of the actual workload and performance on the CYBER 74.

The accounting data proved to be a valuable data base for the CYBER 74 computer system. For purposes of this investigation, the CLARA parameters provided an adequate source of information from which a descriptive workload characterization could be derived. As discussed later, suggested improvements upon the data gathering capabilities could lead to a more comprehensive and detailed workload analysis.

In this thesis, the workload for a number of investigation days were characterized by several measurement tools. These workload characterizations were selected to represent a few of the many variations of workload upon the CYBER 74 computer system, and may be valuable in the initial efforts of a computer performance evaluation. Perhaps more valuable is the fact that this thesis offers for future investigators, a methodology for analyzing a larger representative sample of accounting data.

The effects of variations in workload upon the performance were measured. These measurements can be of value as a preliminary examination of the CYBER 74 and as a guide for developing hypotheses to evaluate computer performance.

One major area where further improvements could be continued involves the enhancement of the reduction capability on the accounting data. Several areas which would benefit from such an effort have been discussed in the Results chapter. In addition, several other recommendations are listed below which would give the accounting data a more complete representation of the actual workload.

Recommendations for Enhancement

The depth of analysis possible from a use of the accounting data would be improved if the problems discussed in the Results chapter were corrected and additional parameters were recorded. A major problem could be eliminated by recording the INPUTQ time in the Execution file [illegible] in the Queue file. This would eliminate the SEARCH module and thus would reduce the processing time of the Executive program by two-thirds.

The resource requests which are stated on each job card should also be recorded in the Execution file. An important consideration for future performance analysis would involve an investigation of the order in which jobs are selected from the input queue, since job selection from the input queue is partially based upon resource requests.

An interview with Mr. Lou Venuti, a member of the Program Control Branch at the ASD Computer Center, indicated that a previous investigation of the amount of swapping in and swapping out had yielded an inordinate amount of data and was unusable for any comprehensive study. Although the accounting system does not presently have the capability, it is suggested that an effort be made toward investigating the swapping.

The CLARA system should include two extra parameters: the amount of time spent swapped out, and the number of times that a job is swapped out. These parameters could possibly be used in an initial investigation of the amount of thrashing or unnecessary swapping in and out that may occur in the CYBER 74.

In order to make the CLARA more easily accessible, the generation of the CLARA magnetic tape files, Queue and Execution files, should be consistently recorded using the same format. Specifically, the end-of-file marks (EOF) between files were found to be inconsistent; in some cases, three EOFs appeared between each day's data, while in other cases only one EOF appeared. This created difficulties in skipping forward to the correct location on a magnetic tape to investigate a particular day's data.

A solution to this problem could be the use of multiple file tapes. These tapes have a label for each file; therefore, each day could be identified by counting records or by referring to it by name.

With the proposed installation of the ECS on the CYBER 74 computer system, the parameters which measure the requests for, and use of, the ECS should also be stored on the CLARA Execution file for future performance analysis investigations.

The accounting system does not presently have the capability to gather data on response time received by interactive terminals. A simple measure, such as average response time per job, would be valuable for future investigations. Without this measure, or similar measures, there is no way to assess time-sharing capabilities.

The addition of a parameter, the time a job leaves the output queue, would allow further investigation into possible bottlenecks

in job throughput.

A final recommendation is that a better or more accurate representation for the distribution of resource usage for the subroutine UNIFORM be developed.

Future Considerations

In this section, possibilities for future use of the workload characterization and the performance measures from the CYBER 74 computer system will be discussed. Several possible applications for these data are listed below (Ref 29:24).

(1) Analyzing the effects of a computer system modification.

(2) Providing data for a performance improvement investigation or effort.

(3) Designing synthetic workloads and workload models for simulation or analytical performance models.

(4) Devising computer charging schemes and algorithms.

(5) Compiling trends of past usage for projecting future demands on the computer system.

(6) Providing management with current workload characteristics and performance measures.

The methodology developed in this thesis may be of value in measuring the effect of a system modification, such as the scheduled addition in 1977 of the Extended Core Storage (ECS) to the CYBER 74. A more extensive data base could be developed which would include an investigation of a larger number of days.

Dr. R. Watson describes several techniques which can be used to determine the effect of a modification (Ref 29:25). Many days of accounting data before and after the modification may be analyzed to measure changes in performance. The initial steps toward this approach

were the same as those completed in this thesis. Weekends, holidays, and days when the computer was down were excluded; the time of day to be investigated was restricted to daytime operation in order to narrow the measures of performance. During the selected time periods, the average values for various performance measures and workload characteristics were calculated; these average values could be used to measure the modification's effect on performance.

An important consideration with this approach is to determine whether the changes in performance after installation of the ECS are caused by fluctuations in the workload characteristics or by the ECS modification. The influence of the workload characteristics upon the performance must be eliminated or significantly reduced.

In recognition of this problem, Watson suggests two independent methods of analyzing the accounting data (Ref 29:34). Both can adequately compare performance measures derived from a computer system by reducing the influence of the varying workloads.

The first method involves use of multiple regression analysis upon the workload characteristics in order to estimate each performance measure. A modification variable is added to the workload characteristics during regression analysis, and a regression coefficient is derived to represent the estimated change resulting from the system modification. In this manner, the variation in workloads can be significantly reduced in order to measure the change in system performance resulting from the modification.

The Cluster Analysis method is a second approach to measuring the effect of a system modification. This method reduces the effect of

workload fluctuations by clustering similar workloads for investigation before and after the modification. The clustering technique suggested by Watson compares the similarity of the workloads with the use of frequency density functions or the average workload characteristics. The results of any performance change is based on the assumption that when no modifications are made, similar workloads should produce similar performance.

Both methods for measuring the effect of a system modification could be used to evaluate the addition of the ECS to the CYBER 74 computer system. This thesis investigation provides workload job data which could be used for the linear regression analysis, as well as probability density functions, average workload characteristics, and average performance measures which could be used as input for the clustering method.

Bibliography

1. Agrawala, A. K., J. M. Mohr, and R. M. Bryant. "An Approach to the Workload Characterization Problem." Computer, 9:18-32 (June 1976).

2. ASD Computer Center. Introduction to ASD Computer Center, Wright-Patterson Air Force Base, Ohio: ASD Computer Center, May 1974.

3. Bard, Y. "Performance Criteria and Measurement for a Time-Sharing System." IBM Systems Journal, 10: 193-216 (1971).

4. Bell, T. E., B. W. Boehm, and R. A. Watson. Computer Performance Analysis: Framework and Initial Phases for a Performance Improvement Effort. Rand Technical Report R-549-1-PR. Santa Monica: Rand Corp., November 1972.

5. Bell, T. E. Computer Performance Analysis: Measurement Objectives and Tools. Rand Technical Report R-584-NASA/PR. Santa Monica: Rand Corp., February 1971.

6. Blitt, William J. et al. The Advanced Logistics System CYBER 73: A Simulation Model. Masters Thesis. Air Force Institute of Technology December 1975. AD-A019841.

7. Conti, D. M. "A Case Study in Monitoring the CDC 6700 - A Multi-Programming, Multi-Processing, Multi-Mode System." Proceedings of Eighth Meeting of CPEUG, 115-118. (4-7 December 1973).

8. Control Data Corporation. Scope Integrated Scheduler Tuning Guide Programming Systems Bulletin. Pub. No. 60493500. Sunnyvale, CA.: CDC, 1975.

9. Control Data Corporation. 6000 Series Introduction and Peripheral Processor Training Manual. Publication Number 60100000. Minneapolis, Minnesota: Control Data Corporation.

10. Dodson, P. O. Control Data Corporation's CYBER 70: Procedures for Performance Evaluation. Masters Thesis. Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio: May, 1974.

11. Drummond, M. E. "A Perspective on System Performance Evaluation." IBM Systems Journal. 8, No. 4: 252-263. (1969).

12. Esposito, J. E. Statistical Analysis to Determine Digital Computer Workload Characteristics. Bedford, Massachusetts: Mitre Corporation, June 1974. AD 786061

13. Esposito, J. E. and K. Sceenivasan. A Semi-Empirical Approach for Evaluating Computer System Performance. ESD-TR-73-297. L. G. Hanscom Field, Massachusetts: Deputy for Command and Management Systems, December 1973. AD 773312.

14. Gray, M. A. "A Performance Study of the Multiprocessor Transition in a Large Computer System." Proceedings of Twelfth Meeting of CPEUG, 33-39. (8-12 November 1976)

15. Gudes, E. and C. Sechler. "Measures for Workload and their Relation to Performance Evaluation." Proceedings of CPEUG: 115-121 (23-26 September 1975).

16. Hamlet, R. G. "Efficient Multiprogramming Resource Allocation and Accounting." Communications of the ACM, 16:337-343 (June 1973).

17. Hellerman, Herbert and Thomas F. Conroy. Computer System Performance. New York: McGraw-Hill Book Company, 1975.

18. Johnson, R. R. "Needed: A Measure for Measure." Datamation, 22-30. (15 December 1970).

19. Kimbleton, S. R. The Role of Computer System Models in Performance Evaluation. TR-72-4. Ann Arbor: The University of Michigan, Michigan, Department of Industrial Engineering, 30 April 1972. AD 746881.

20. Lucas, H. C., Jr. "Performance Evaluation and Monitoring." Computing Surveys, 3: 79-91 (September 1971).

21. Nie, Norman H., et al. Statistical Package for the Social Sciences. New York: McGraw-Hill Book Company, Inc., 1975.

22. Noe, J. D. and G. J. Nutt. "Validation of a Trace-Driven CDC 6400 Simulation." Proceedings AFIPS Spring Joint Computer Conference, 40:749-758. Montvale, N. J.: AFIPS Press, 1972.

23. Pleuler, E. F. Design of a Computer System Performance Measurement Program. Masters Thesis. Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio: December 1974.

24. Shelter, A. C. and T. E. Bell. Computer Performance Analysis: Controlled Testing. Rand Technical Report R-1436-DCA. Santa Monica: Rand Corp., April 1974.

25. Sreenivasan, K. On the Use of Accounting Data in the Evaluation of Computer System Performance. Bedford, Mass.: Mitre Corp., January 1974.

26. Screenivasan, K. and J. E. Esposito. Experiments with the Burroughs B3500 Computer Using a Synthetic Workload. Bedford, Mass.: Mitre Corp., January 1975. AD/A 004804.

27. Screenivasan, K., et al. Construction and Application of Representative Synthetic Workloads. Bedford, Mass.: Mitre Corp., August 1973. AD 769865.

28. Svobodova, L. <u>Computer Performance Measurement and Evaluation Methods: Analysis and Applications</u>. New York: American Elsevier Publishing Company, Inc., 1976.

29. Watson, R. A. <u>Computer Performance Analysis: Applications of Accounting Data</u>. Rand Technical Report R-573-NASA/PR. Santa Monica: Rand Corp., May 1971.

30. Wood, D. C. and E. H. Forman. "Throughput Measurement Using a Synthetic Job Stream." <u>AFIPS Conference Proceedings</u>. <u>39</u>: 51-56 (November 1971).

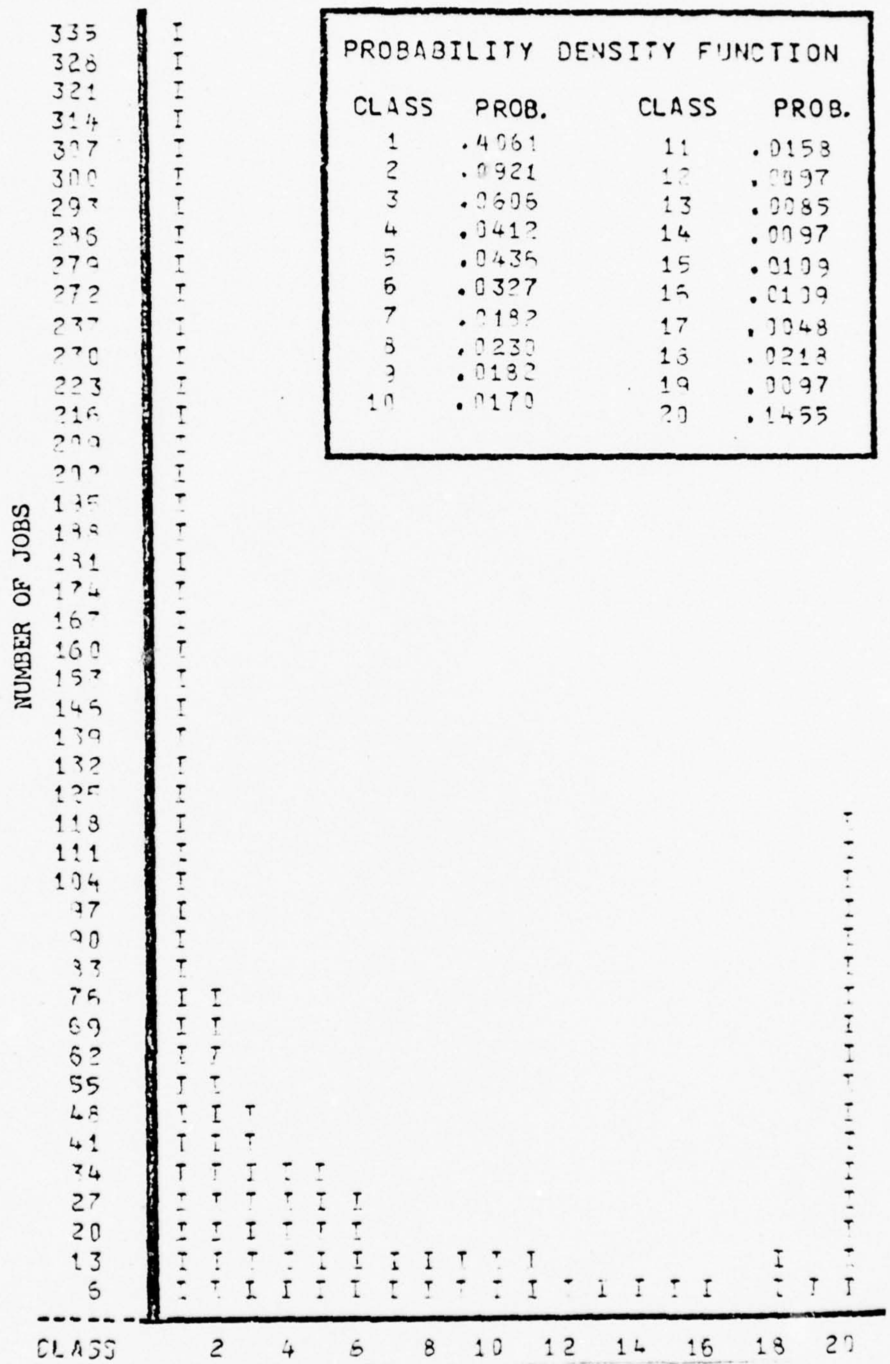
Appendix A

Workload Characterization,

20 October

The following variables are characterized with a frequency histogram, probability density function, and/or a statistical summary. All measurements of time are in seconds.

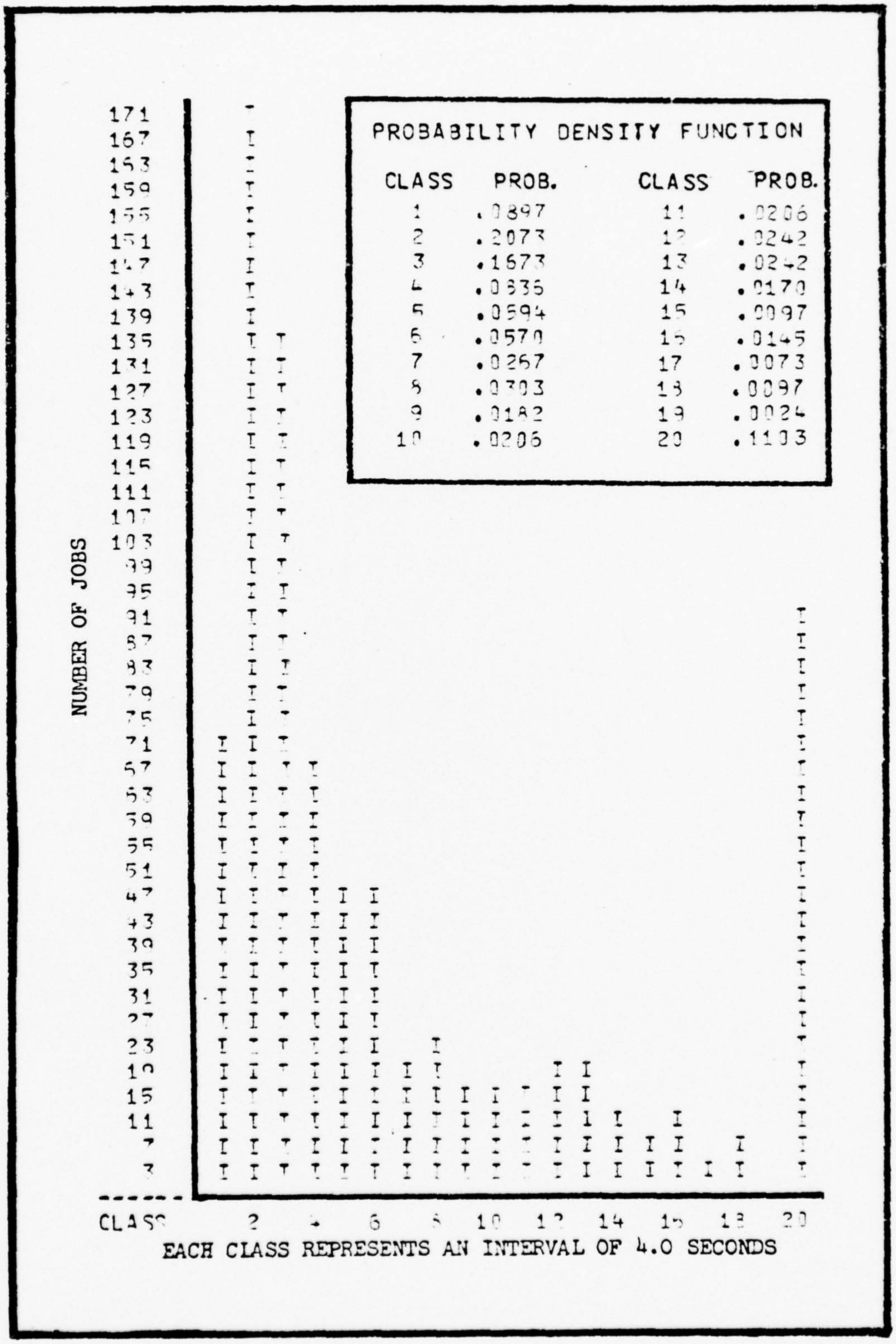
Name	Description
CPTIME	The central processor time needed to process a job.
PPTIME	The peripheral processor time needed to process a job.
TIMEIO	The total I/O time needed to process a job.
TOTCOST	The total job cost in terms of CRUs.
KSW	The memory usage in terms of kilo-word seconds.
CPCT	The total control point occupancy time which is the time a job enters a control point until it leaves a control point for the last time.
CMLOC	The number of central memory locations occupied by all jobs including this job at the time this job left a control point.
CPLOC	The number of control points occupied by all jobs including this job at the time this job left a control point.
ROLLOUT	The total time that a batch job was rolled out for processing magnetic tapes.
IATIME	The total time between this batch job and the last batch job to arrive into the input queue.
INQTIME	The time a batch job spends in the input queue.
TAPEREQ/TAPEUSED	The number of tapes requested/used per job.



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.4061	11	.0158
2	.0921	12	.0097
3	.0606	13	.0085
4	.0412	14	.0097
5	.0436	15	.0109
6	.0327	16	.0109
7	.0182	17	.0048
8	.0230	18	.0218
9	.0182	19	.0097
10	.0170	20	.1455

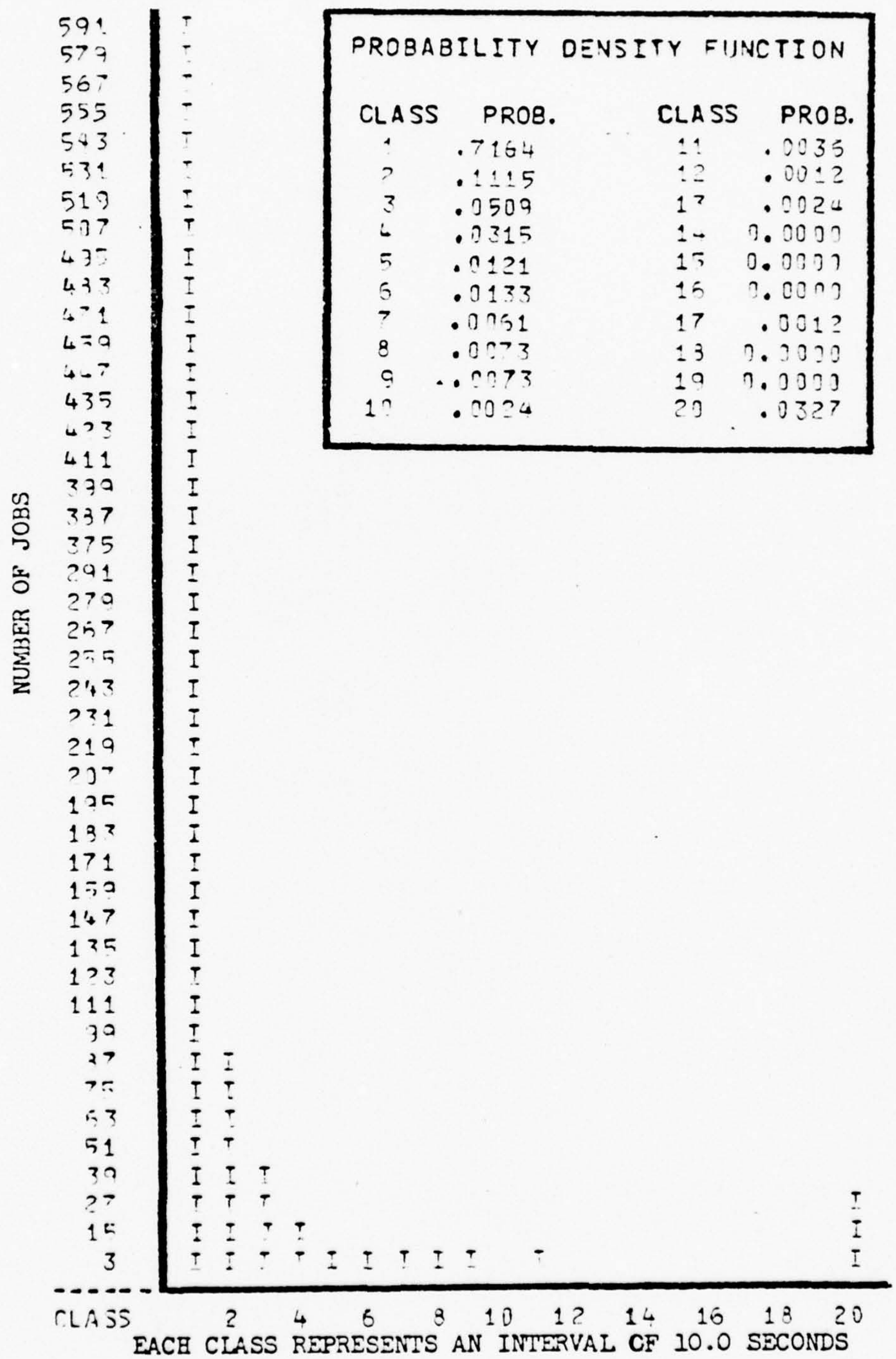
Fig. 30A. Distribution of CPU time for 20 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.0897	11	.0206
2	.2073	12	.0242
3	.1673	13	.0242
4	.0835	14	.0170
5	.0594	15	.0097
6	.0570	16	.0145
7	.0267	17	.0073
8	.0303	18	.0097
9	.0182	19	.0024
10	.0206	20	.1103

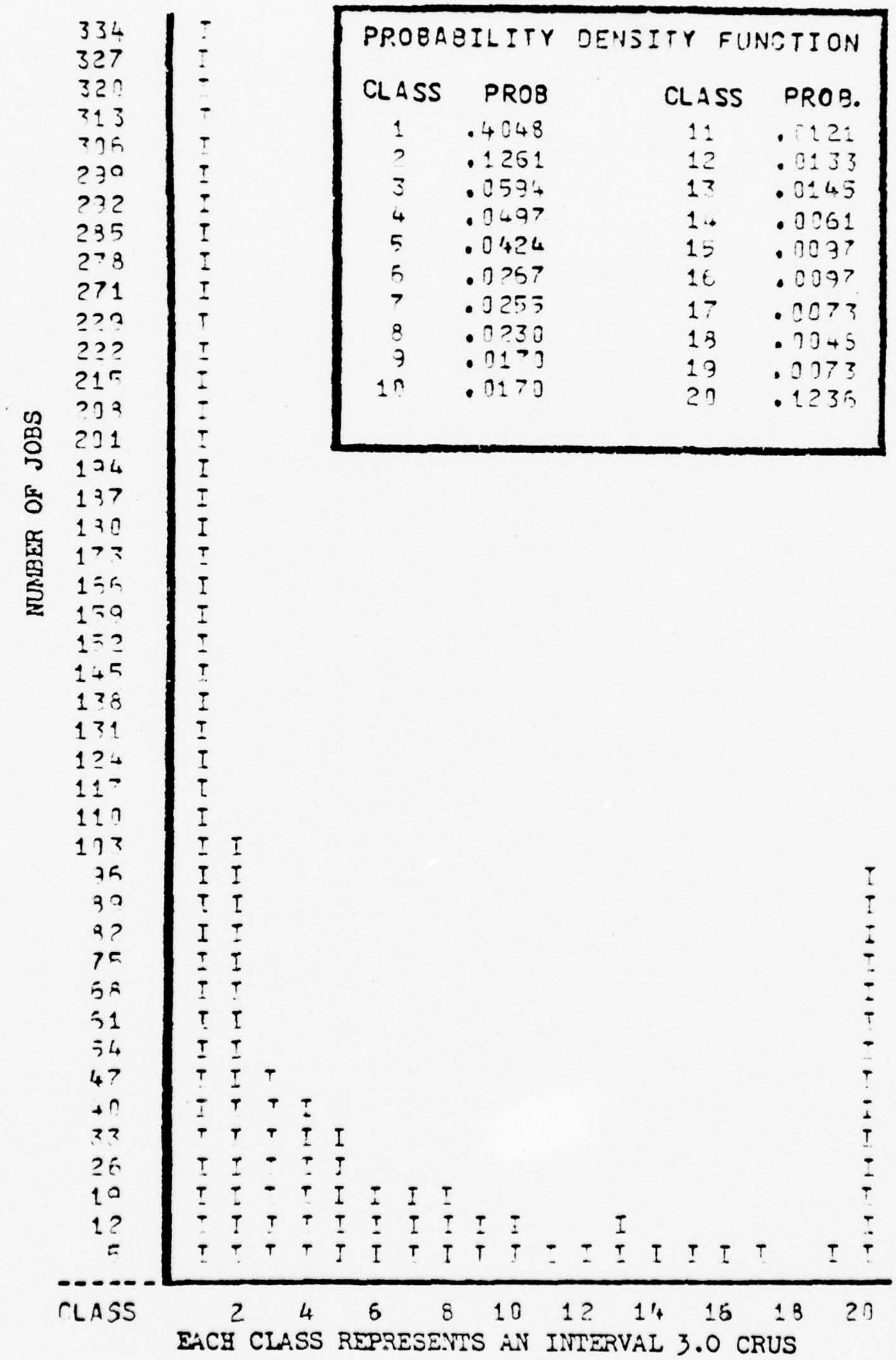
Fig. 30B. Distribution of PPU time for 20 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.7164	11	.0036
2	.1115	12	.0012
3	.0509	13	.0024
4	.0315	14	0.0000
5	.0121	15	0.0000
6	.0133	16	0.0000
7	.0061	17	.0012
8	.0073	18	0.0000
9	.0073	19	0.0000
10	.0024	20	.0327

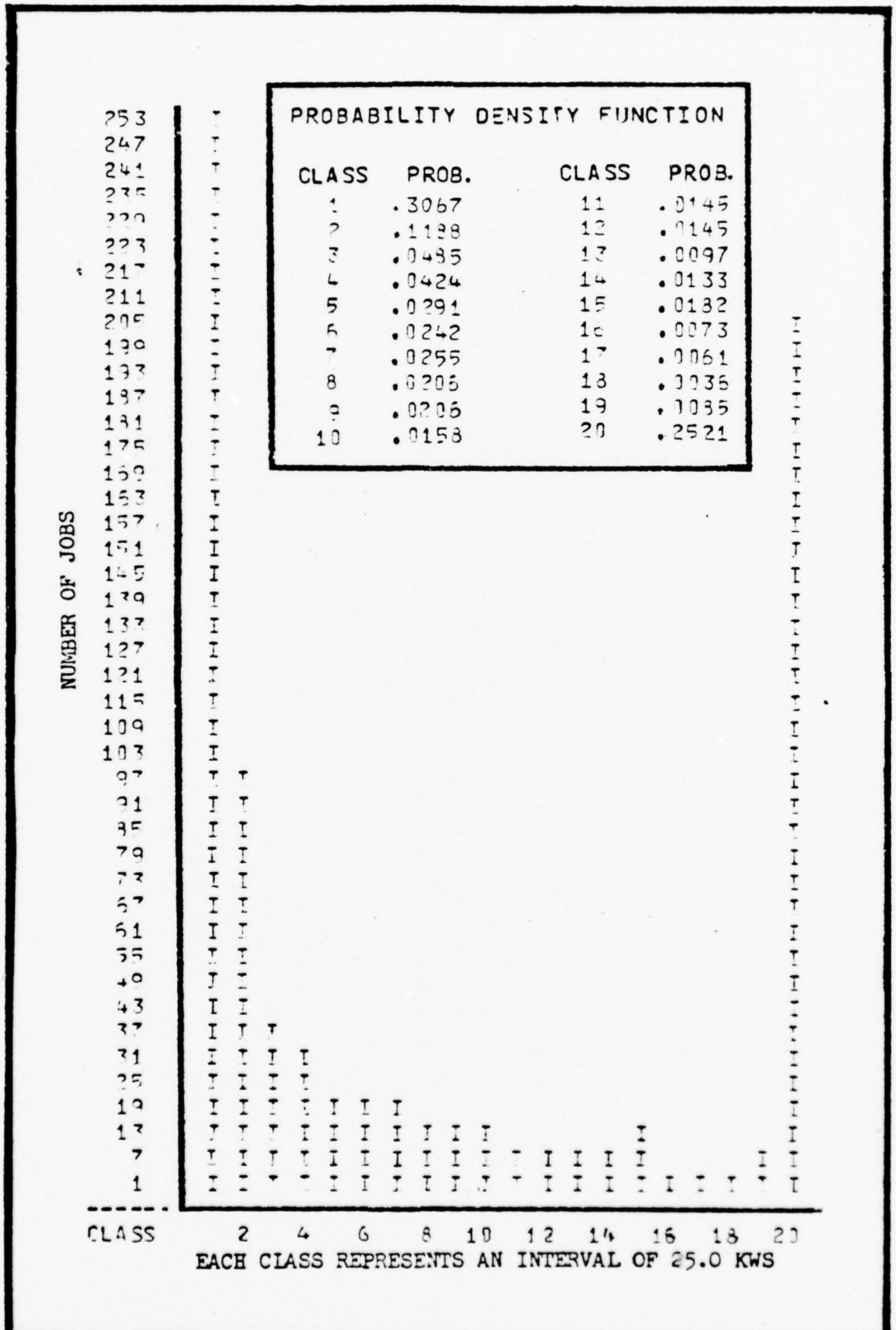
Fig. 30C. Distribution of I/O time for 20 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.4048	11	.0121
2	.1261	12	.0133
3	.0594	13	.0145
4	.0497	14	.0061
5	.0424	15	.0037
6	.0267	16	.0097
7	.0255	17	.0073
8	.0230	18	.0046
9	.0170	19	.0073
10	.0170	20	.1236

Fig. 30D. Distribution of CRUs for 20 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.3067	11	.0145
2	.1188	12	.0145
3	.0485	13	.0097
4	.0424	14	.0133
5	.0291	15	.0182
6	.0242	16	.0073
7	.0255	17	.0061
8	.0206	18	.0035
9	.0206	19	.0085
10	.0158	20	.2521

Fig. 30E. Distribution of KWS for 20 October

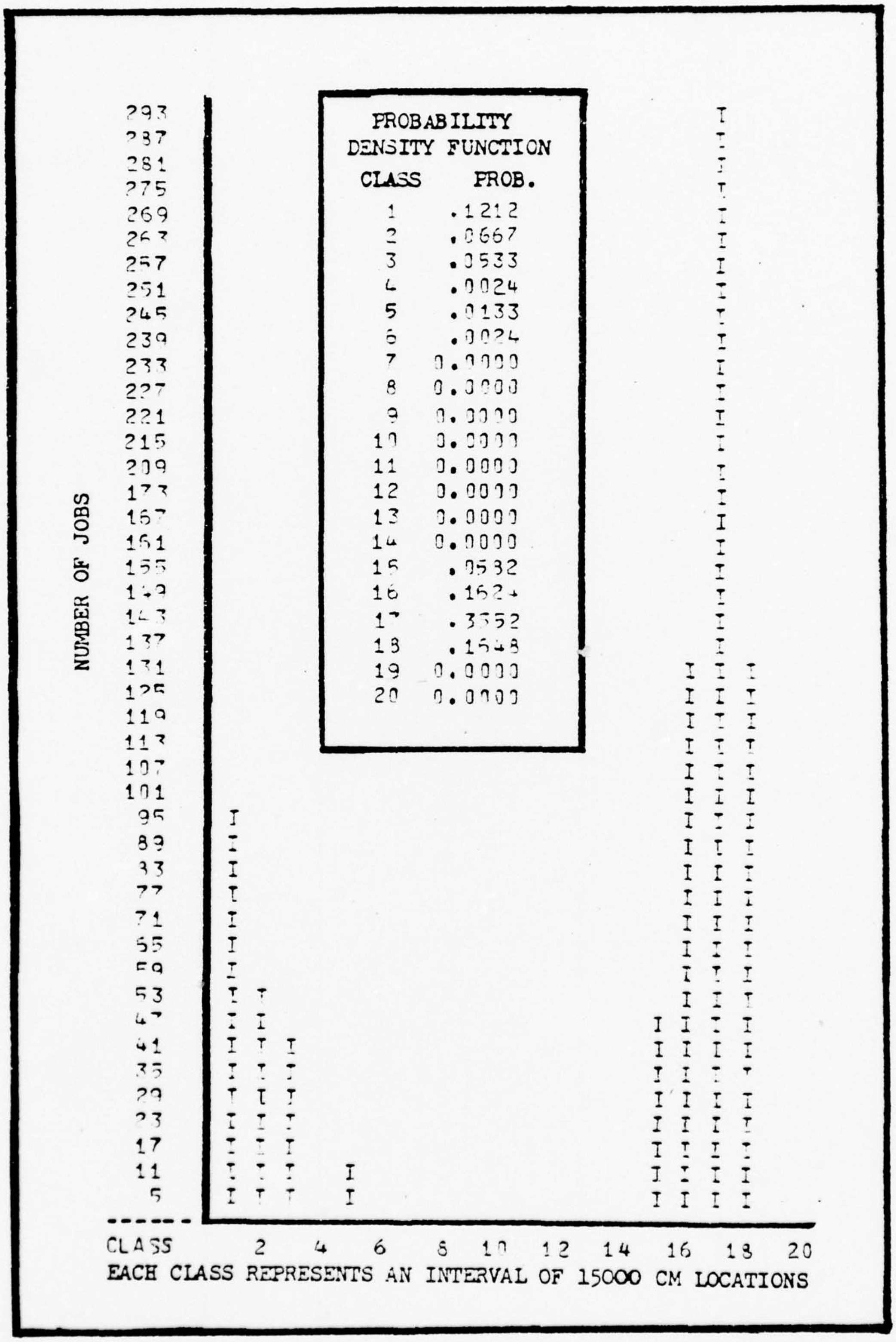


Fig. 30F. Distribution of Central Memory Locations for 20 October

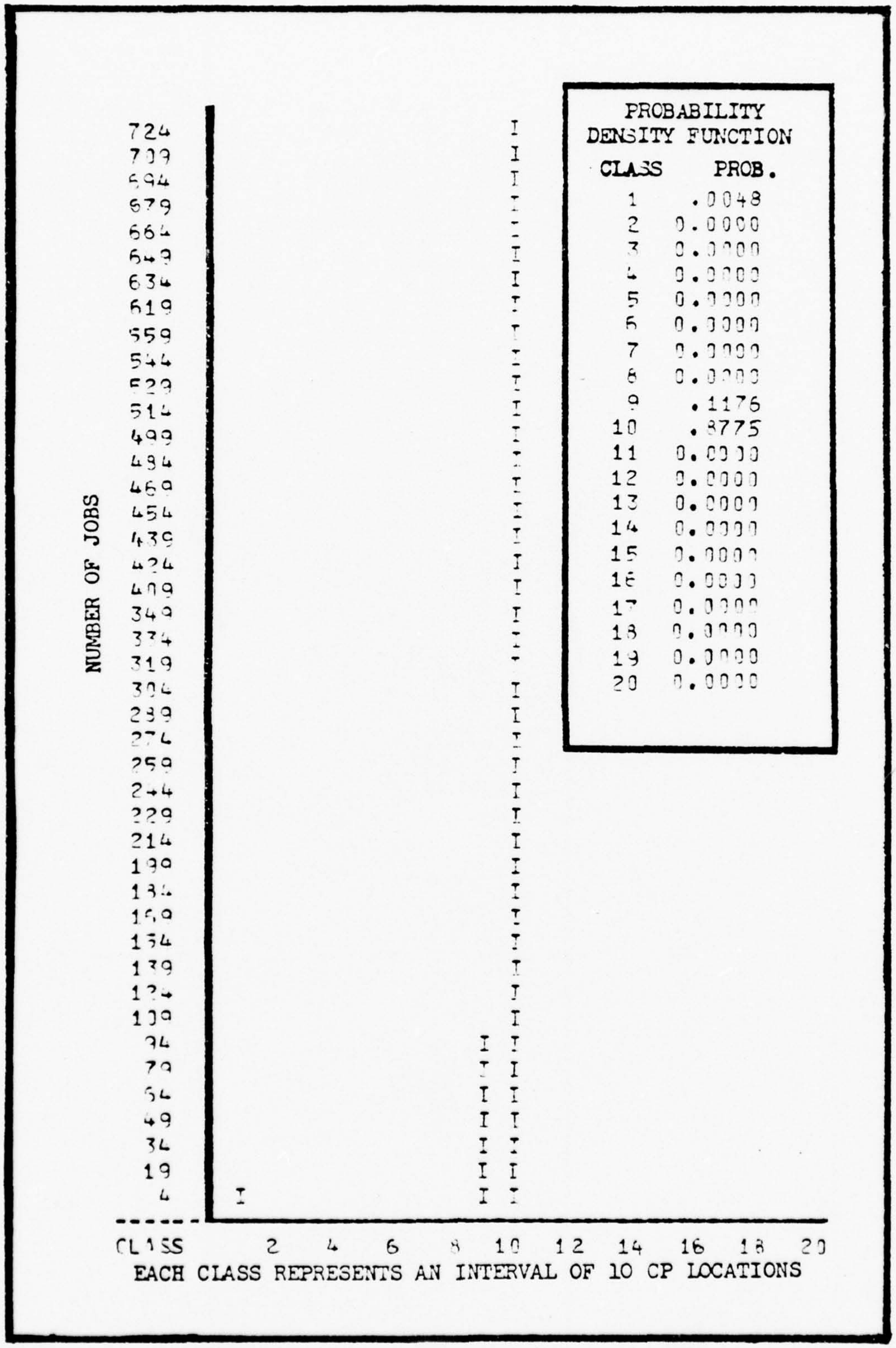
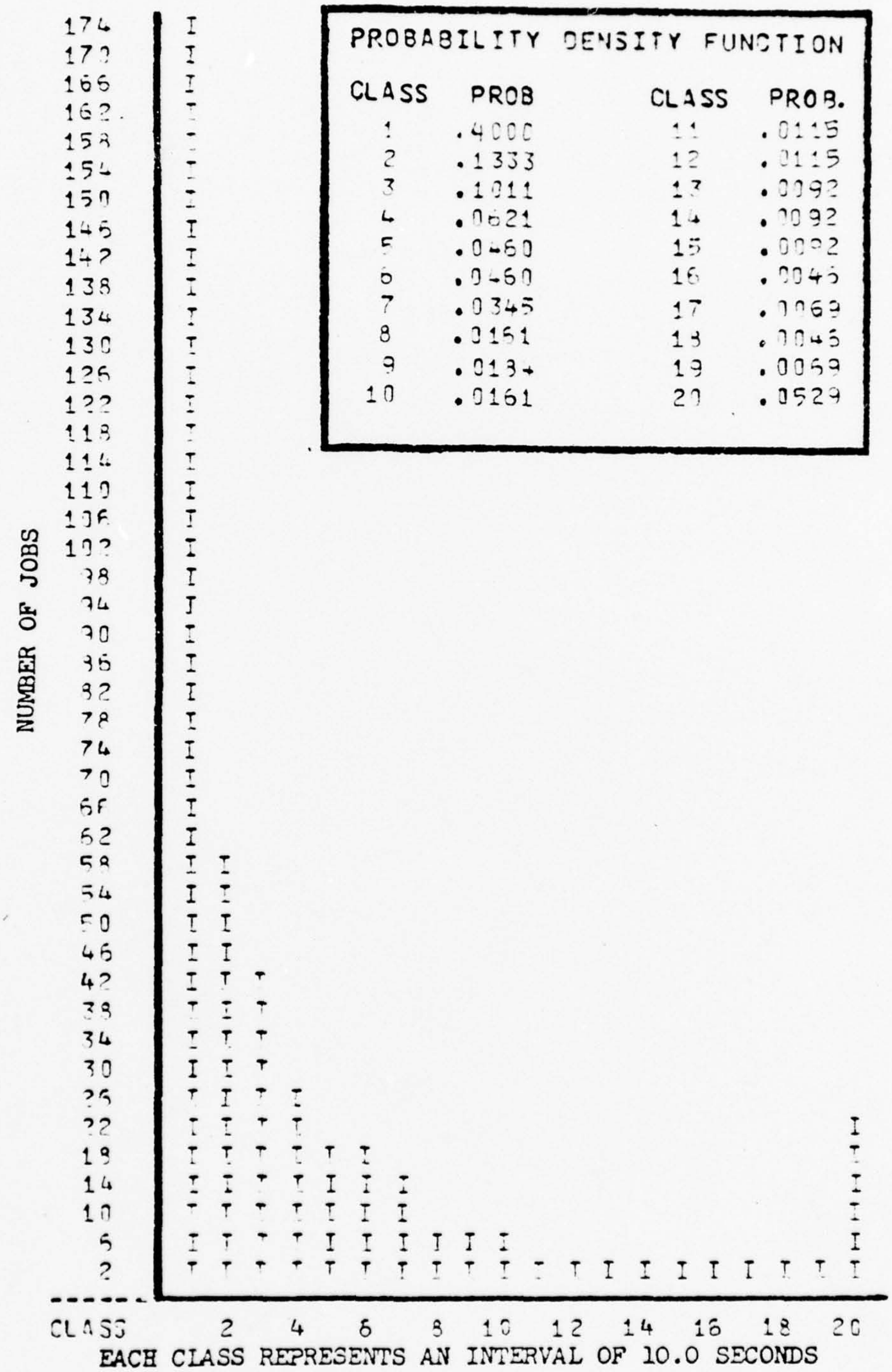


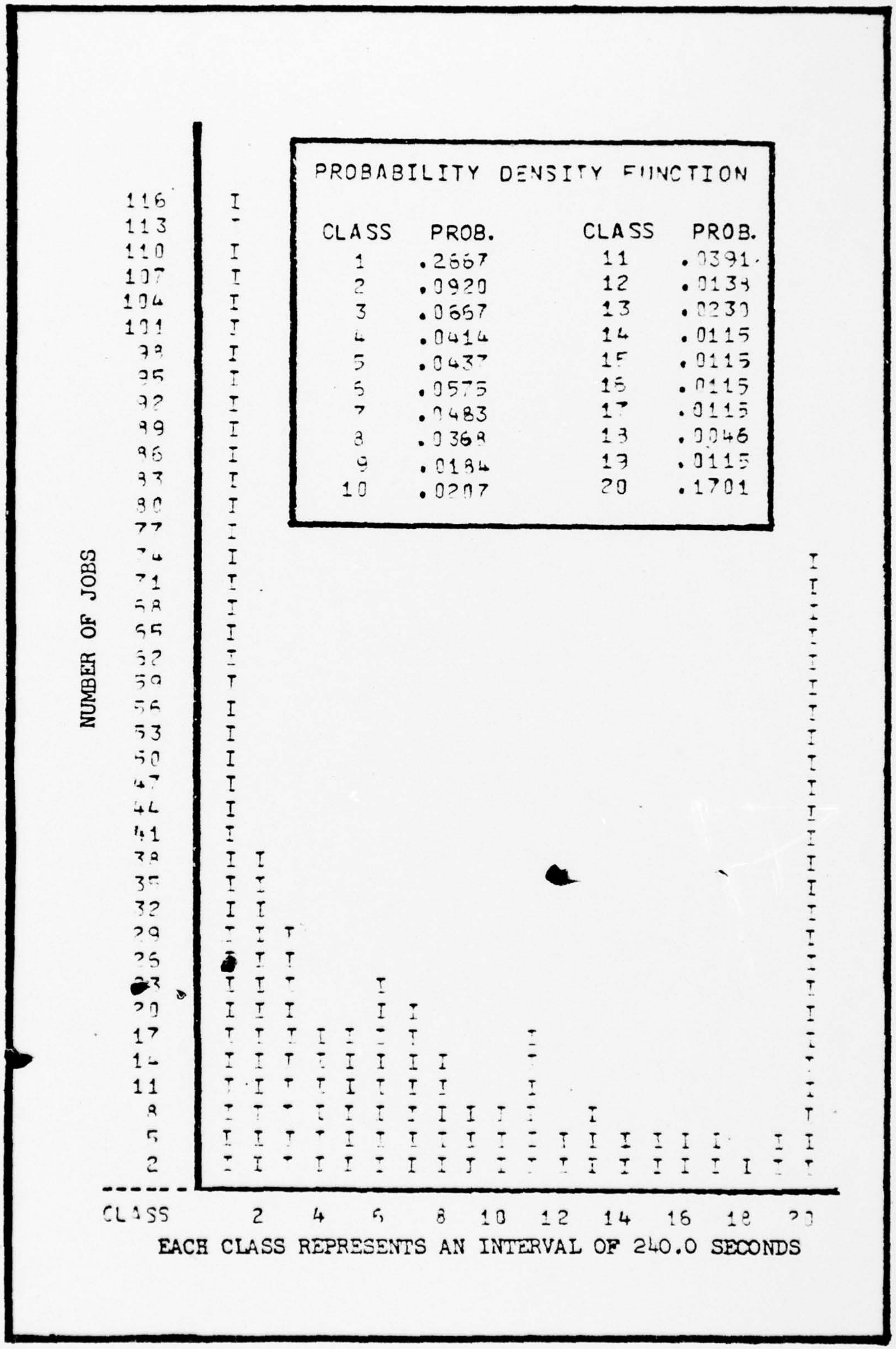
Fig. 30G. Distribution of Control Point Locations for 20 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.4000	11	.0115
2	.1333	12	.0115
3	.1011	13	.0092
4	.0621	14	.0092
5	.0460	15	.0092
6	.0460	16	.0046
7	.0345	17	.0069
8	.0161	18	.0046
9	.0184	19	.0069
10	.0161	20	.0529

Fig. 30H. Distribution of Inter-Arrival time for 20 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.2667	11	.0391
2	.0920	12	.0138
3	.0667	13	.0230
4	.0414	14	.0115
5	.0437	15	.0115
6	.0575	16	.0115
7	.0483	17	.0115
8	.0368	18	.0046
9	.0184	19	.0115
10	.0207	20	.1701

Fig. 30I. Distribution of Input Queue time for 20 October

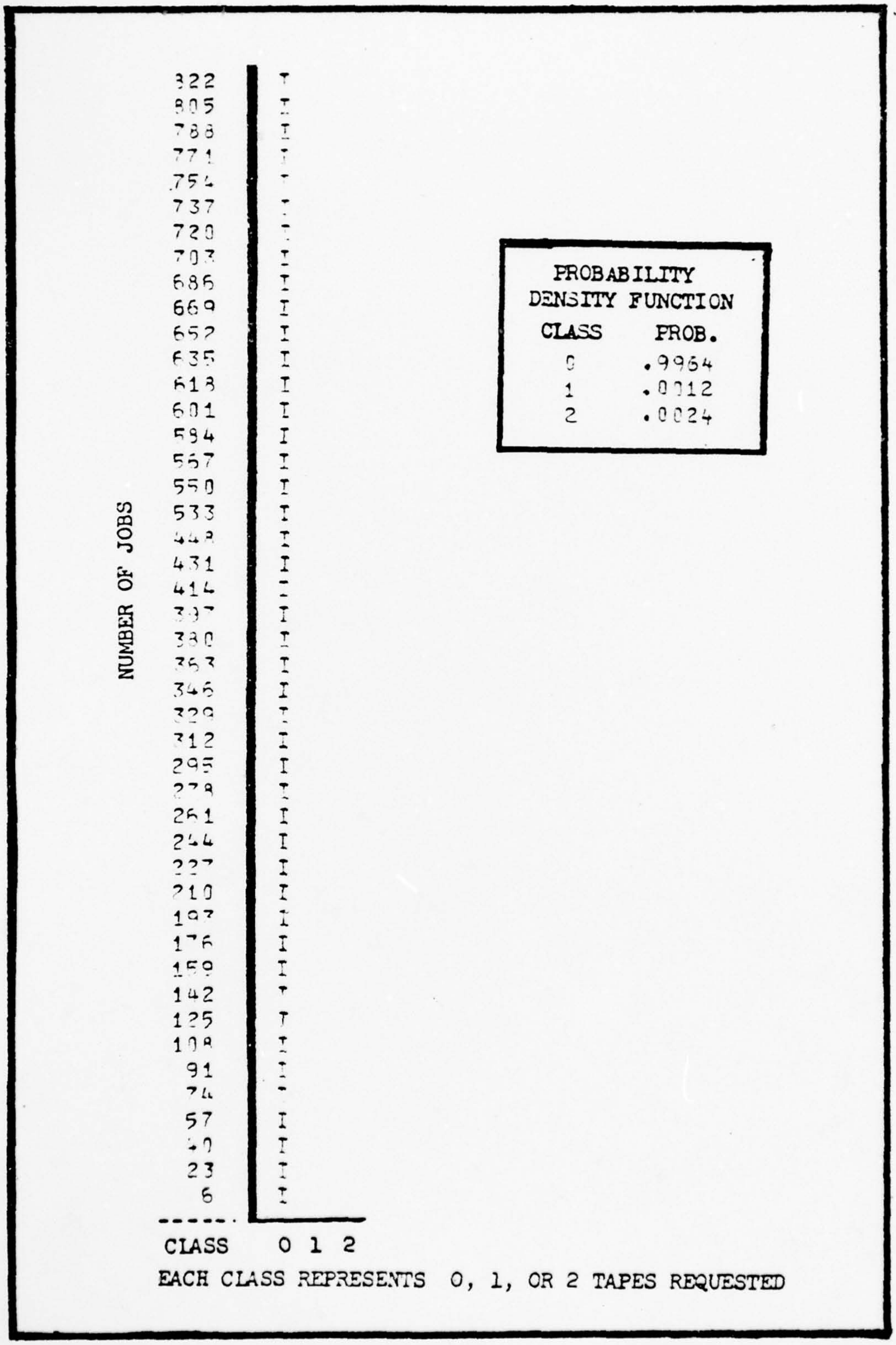


Fig. 30J. Distribution of Tapes Requested for 20 October

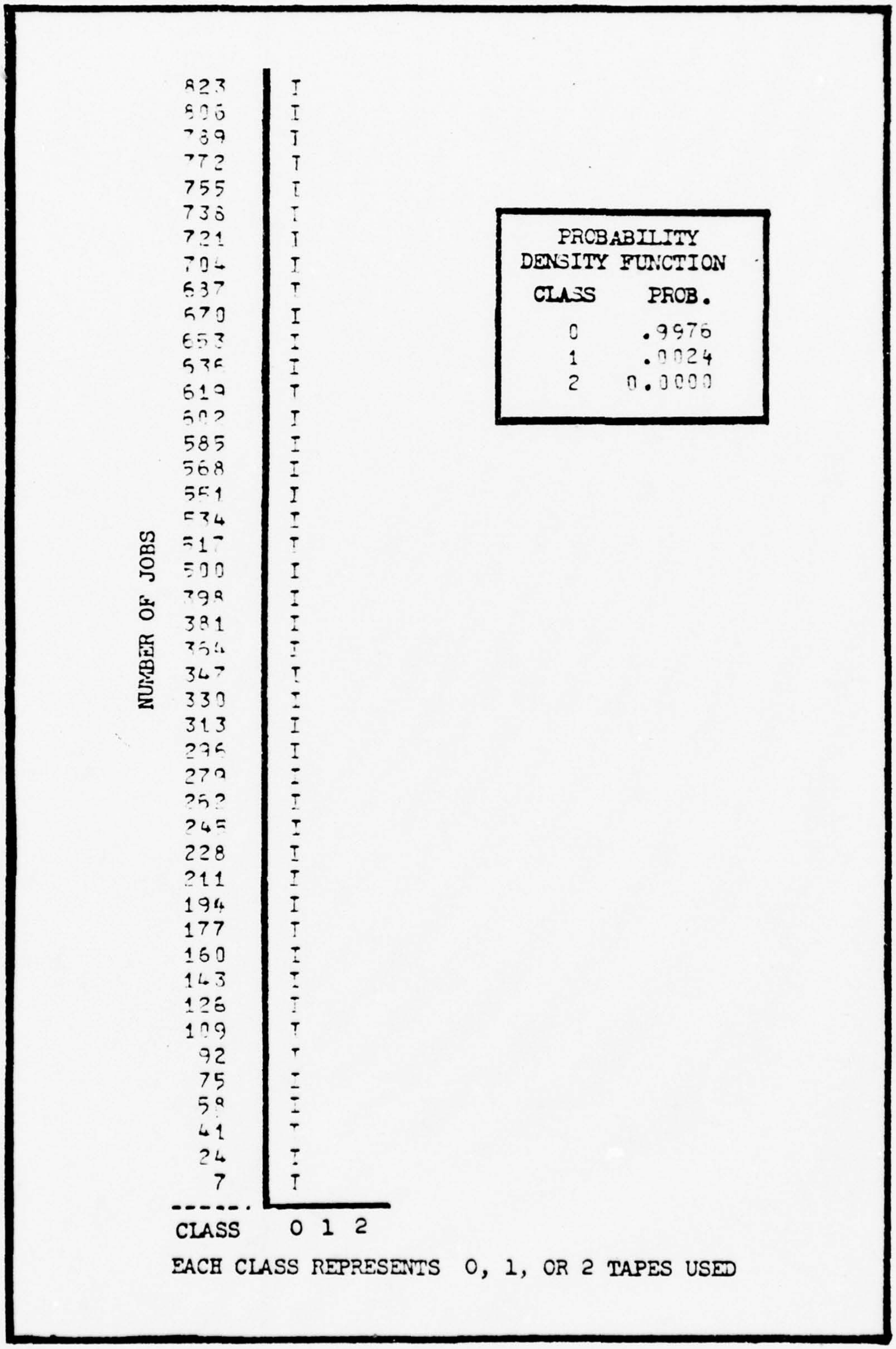


Fig. 30K. Distribution of Tapes Used for 20 October

TABLE XV

Statistical Summary for Workload Parameters,

20 October

Variable CPTIME CPU time in seconds

Mean	22.129	Std Err	1.714	Std Dev	49.221
Variance	2422.702	Kurtosis	31.997	Skewness	4.759
Minimum	.001	Maximum	509.668	Sum	18256.771
C.V. Pct	222.423	.95 C.I.	18.766	to	25.493

Variable PPTIME Peripheral Processor time in seconds

Mean	48.449	Std Err	8.657	Std Dev	248.652
Variance	61828.048	Kurtosis	580.836	Skewness	22.725
Minimum	.281	Maximum	6580.852	Sum	39970.335
C.V. Pct	513.226	.95 C.I.	31.457	to	65.441

Variable TIMEIO Input-Output time in seconds

Mean	24.023	Std Err	3.360	Std Dev	96.510
Variance	9314.141	Kurtosis	112.406	Skewness	9.527
Minimum	.037	Maximum	1477.868	Sum	19819.240
C.V. Pct	401.734	.95 C.I.	17.428	to	30.619

Variable TOTCOST Total cost in CRUs

Mean	30.995	Std Err	3.302	Std Dev	94.834
Variance	8993.412	Kurtosis	88.277	Skewness	8.349
Minimum	.040	Maximum	1311.828	Sum	25570.709
C.V. Pct	305.966	.95 C.I.	24.514	to	37.475

TABLE XV (continued)

Variable KWS Memory in kilo-word seconds

Mean	834.037	Std Err	101.257	Std Dev	2908.388
Variance	8458718.573	Kurtosis	81.254	Skewness	8.256
Minimum	0.000	Maximum	37101.740	Sum	688080.388
C.V. Pct	348.712	.95 C.I.	635.285	to	1032.789

Variable CPCT Control Point Occupancy Time in seconds

Mean	1080.859	Std Err	72.719	Std Dev	2088.689
Variance	4362622.437	Kurtosis	24.090	Skewness	4.149
Minimum	1.000	Maximum	18637.000	Sum	891709.000
C.V. Pct	193.243	.95 C.I.	938.123	to	1223.595

Variable CMLCC Central Memory Locations in Use

Mean	187128.242	Std Err	3431.430	Std Dev	98560.329
Variance	9714E+10	Kurtosis	-.726	Skewness	-1.093
Minimum	512.000	Maximum	261632.000	Sum	.1543E+09
C.V. Pct	52.670	.95 C.I.	180392.869	to	193863.615

Variable CPLCC Control Points in Use

Mean	95.439	Std Err	.256	Std Dev	7.320
Variance	53.577	Kurtosis	99.653	Skewness	-8.594
Minimum	7.000	Maximum	100.000	Sum	78737.000
C.V. Pct	7.669	.95 C.I.	94.939	to 95.939	

Variable ROLLOUT Total Time Rolled Out in seconds

Mean	22.331	Std Err	6.456	Std Dev	185.437
Variance	34387.057	Kurtosis	200.061	Skewness	13.167
Minimum	0.000	Maximum	3423.000	Sum	18423.000
C.V. Pct	830.407	.95 C.I.	9.659	to	35.003

TABLE XV (continued)

Variable IATIME Inter-arrival time in seconds

Mean	46.038	Std Err	3.742	Std Dev	78.680
Variance	6190.581	Kurtosis	15.959	Skewness	3.525
Minimum	0.000	Maximum	658.000	Sum	20349.000
C.V. Pct	170.901	.95 C.I.	38.683	to	53.394

Variable INQTIME Time in Input Queue in seconds

Mean	1651.144	Std Err	143.952	Std Dev	4015.205
Variance	1612E+08	Kurtosis	15.534	Skewness	3.804
Minimum	0.000	Maximum	23609.000	Sum	1284668.000
C.V. Pct	243.162	.95 C.I.	1368.663	to	1933.825

Variable TAPEREQ Number of Tapes Requested

Mean	.057	Std Err	.009	Std Dev	.266
Variance	.071	Kurtosis	44.331	Skewness	5.861
Minimum	0.000	Maximum	3.000	Sum	47.000
C.V. Pct	466.993	.95 C.I.	.039	to	.075

Variable TAPEDR Number of Tapes Used

Mean	.053	Std Err	.008	Std Dev	.235
Variance	.055	Kurtosis	21.087	Skewness	4.523
Minimum	0.000	Maximum	2.000	Sum	44.000
C.V. Pct	441.241	.95 C.I.	.037	to	.059

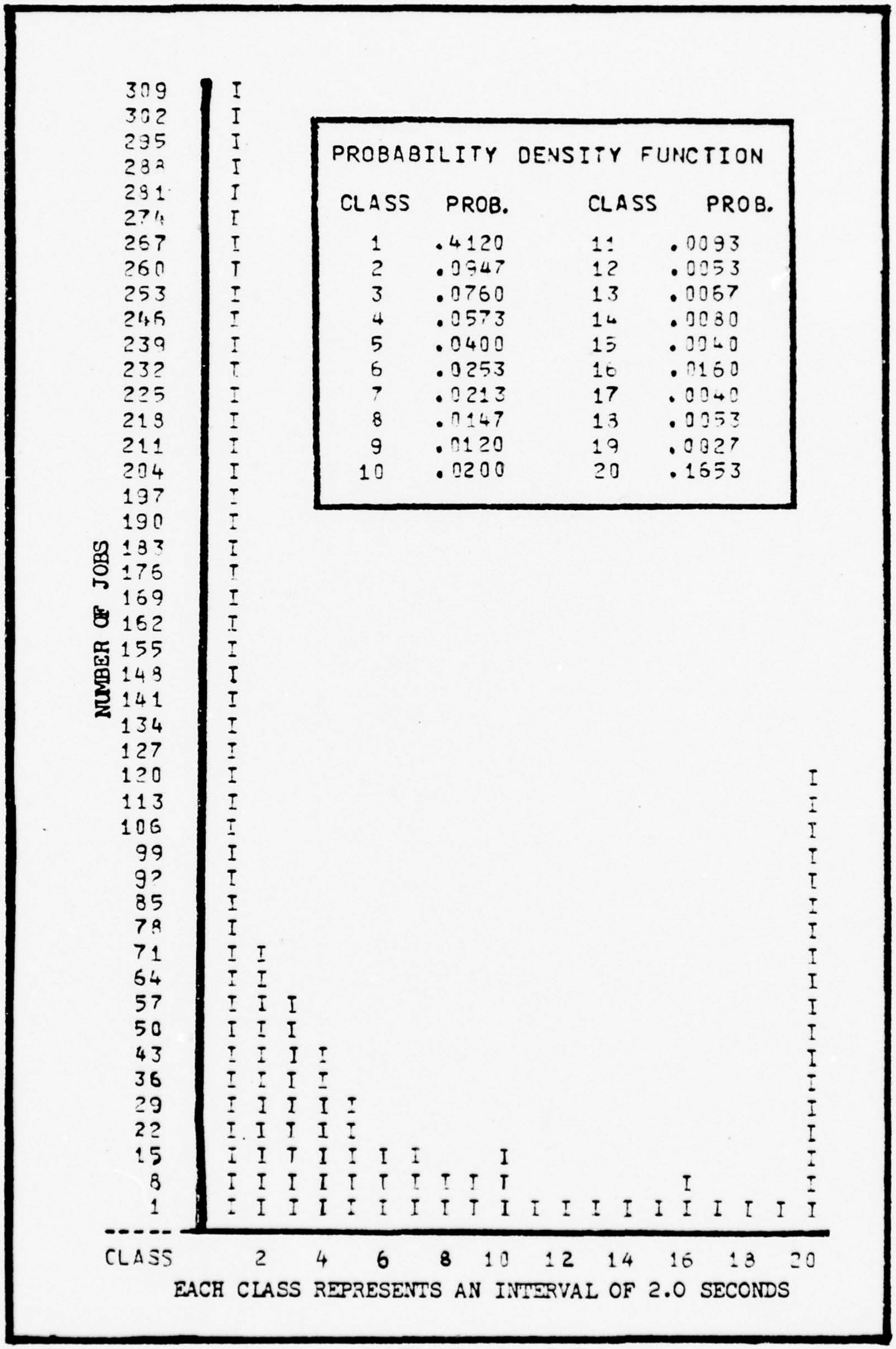
Appendix B

Workload Characterization,

7 October

The following variables are characterized with a frequency histogram, probability density function, and/or a statistical summary. All measurements of time are in seconds.

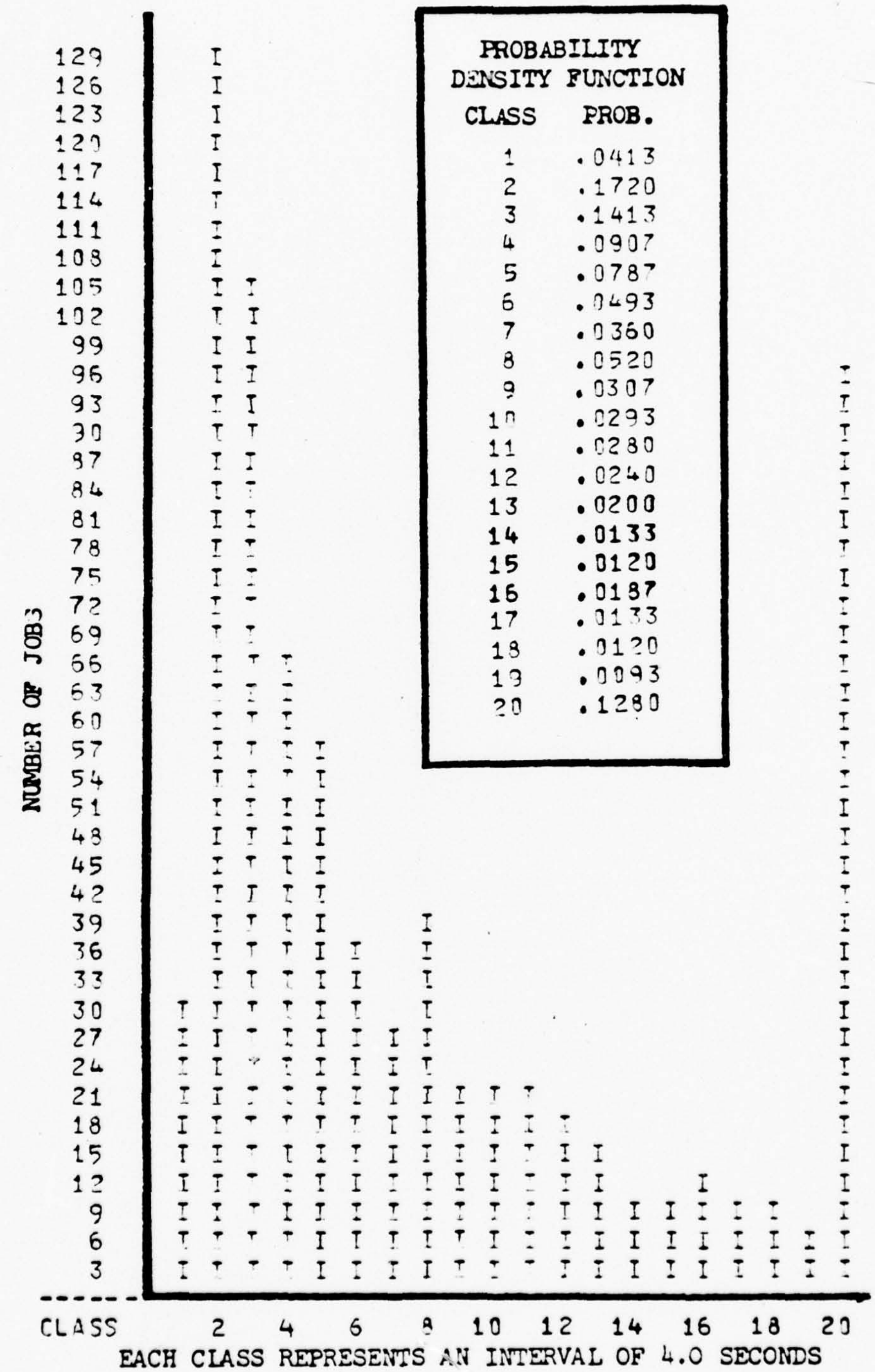
Name	Description
CPTIME	The central processor time needed to process a job.
PPTIME	The peripheral processor time needed to process a job.
TIMEIO	The total I/O time needed to process a job.
TOTCOST	The total job cost in terms of CRUs.
KSW	The memory usage in terms of kilo-word seconds.
CPOT	The total control point occupancy time which is the time a job enters a control point until it leaves a control point for the last time.
CMLOC	The number of central memory locations occupied by all jobs including this job at the time this job left a control point.
CPLOC	The number of control points occupied by all jobs including this job at the time this job left a control point.
ROLLOUT	The total time that a batch job was rolled out for processing magnetic tapes.
IATIME	The total time between this batch job and the last batch job to arrive into the input queue.
INQTIME	The time a batch job spends in the input queue.
TAPEREQ/TAPEUSED	The number of tapes requested/used per job.



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.4120	11	.0093
2	.0947	12	.0053
3	.0760	13	.0067
4	.0573	14	.0080
5	.0400	15	.0040
6	.0253	16	.0160
7	.0213	17	.0040
8	.0147	18	.0053
9	.0120	19	.0027
10	.0200	20	.1653

Fig. 31A. Distribution of CPU time for 7 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.
1	.0413
2	.1720
3	.1413
4	.0907
5	.0787
6	.0493
7	.0360
8	.0520
9	.0307
10	.0293
11	.0280
12	.0240
13	.0200
14	.0133
15	.0120
16	.0187
17	.0133
18	.0120
19	.0093
20	.1280

Fig. 31B. Distribution of PPU time for 7 October

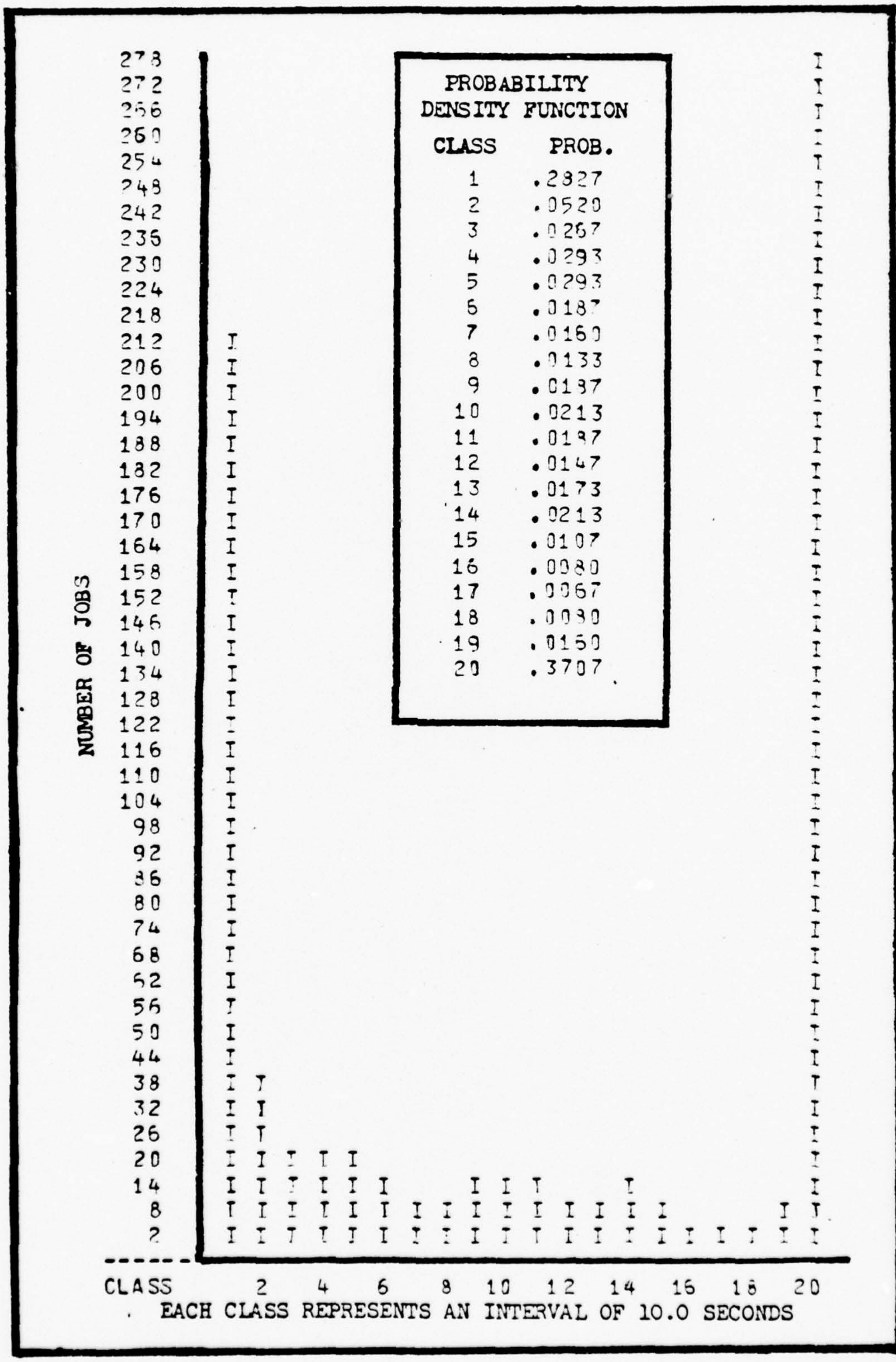


Fig. 31C. Distribution of I/O time for 7 October

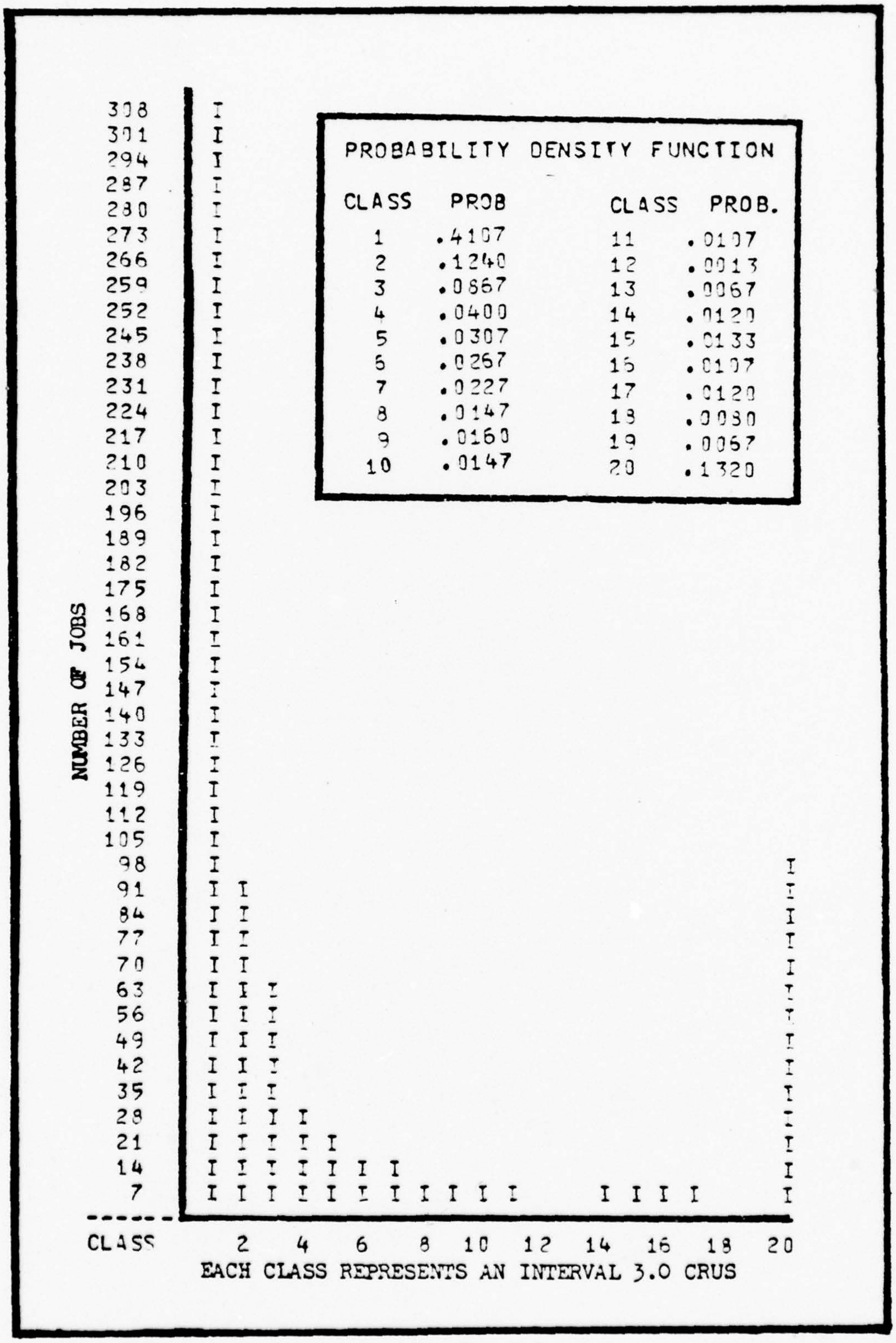
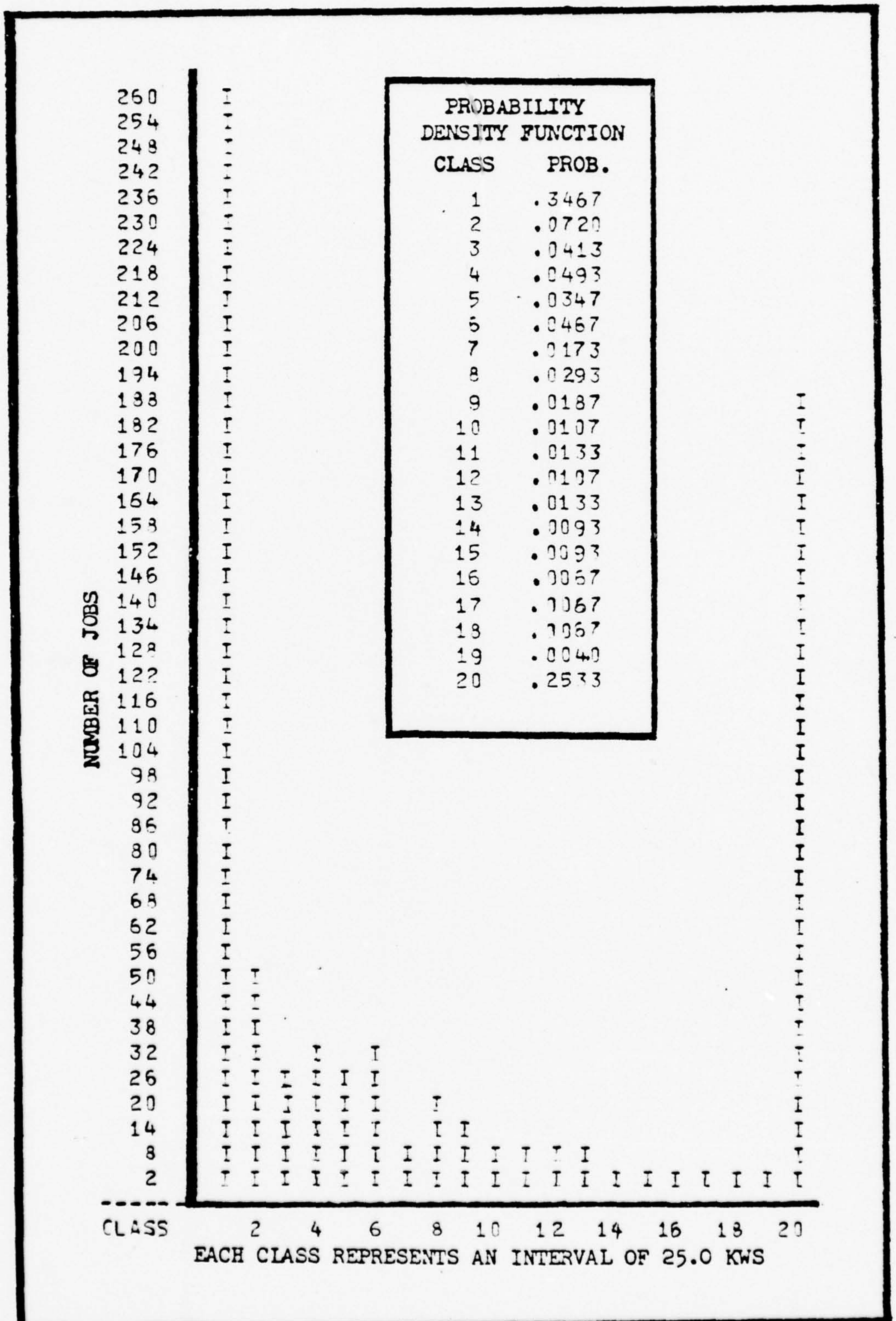


Fig. 31D. Distribution of CRUs for 7 October



CLASS	PROB.
1	.3467
2	.0720
3	.0413
4	.0493
5	.0347
6	.0467
7	.0173
8	.0293
9	.0187
10	.0107
11	.0133
12	.0107
13	.0133
14	.0093
15	.0093
16	.0067
17	.0067
18	.0067
19	.0040
20	.2533

Fig. 31E. Distribution of KWS for 7 October

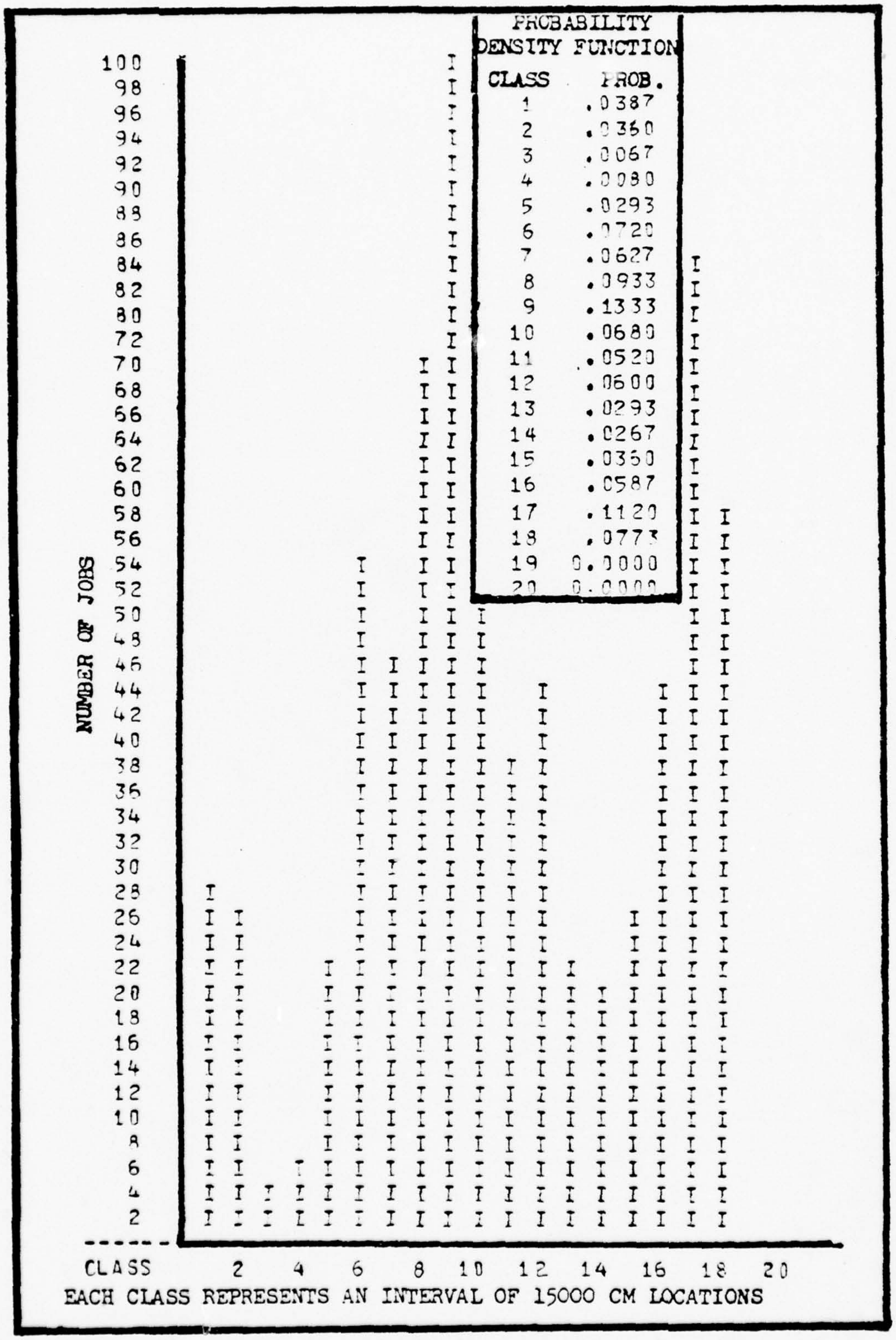
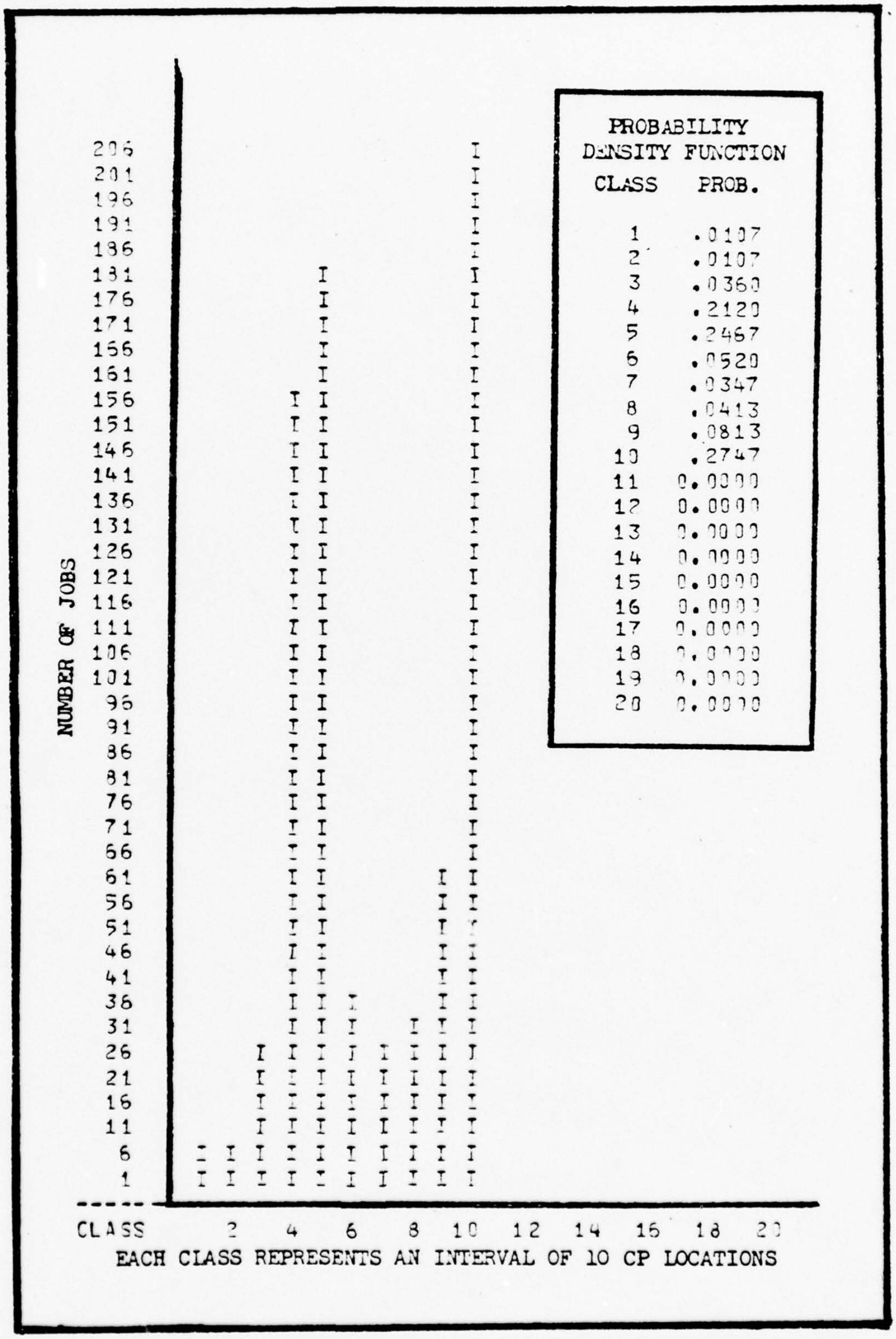


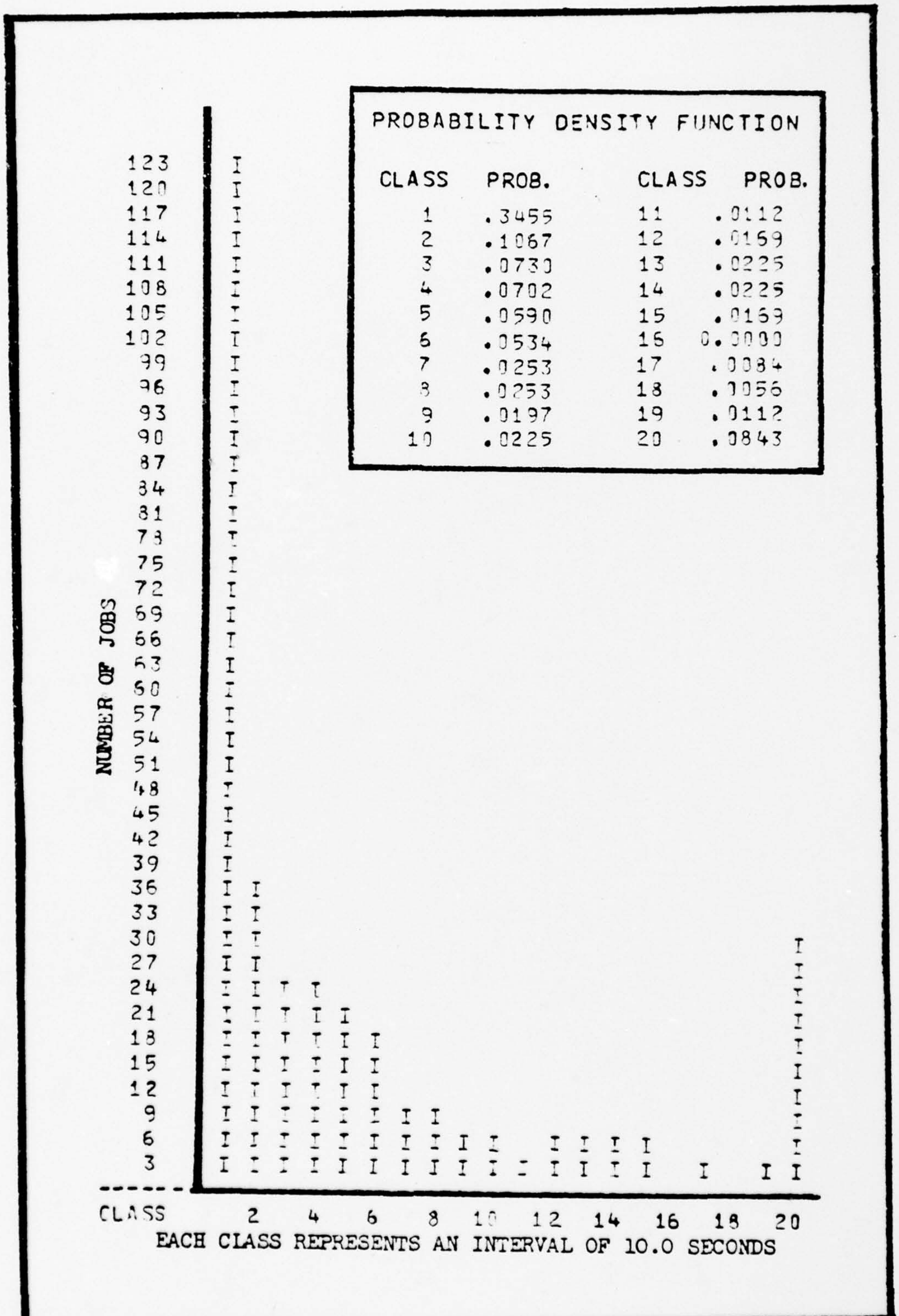
Fig. 31F. Distribution of Central Memory Locations for 7 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.
1	.0107
2	.0107
3	.0360
4	.2120
5	.2467
6	.0520
7	.0347
8	.0413
9	.0813
10	.2747
11	0.0000
12	0.0000
13	0.0000
14	0.0000
15	0.0000
16	0.0000
17	0.0000
18	0.0000
19	0.0000
20	0.0000

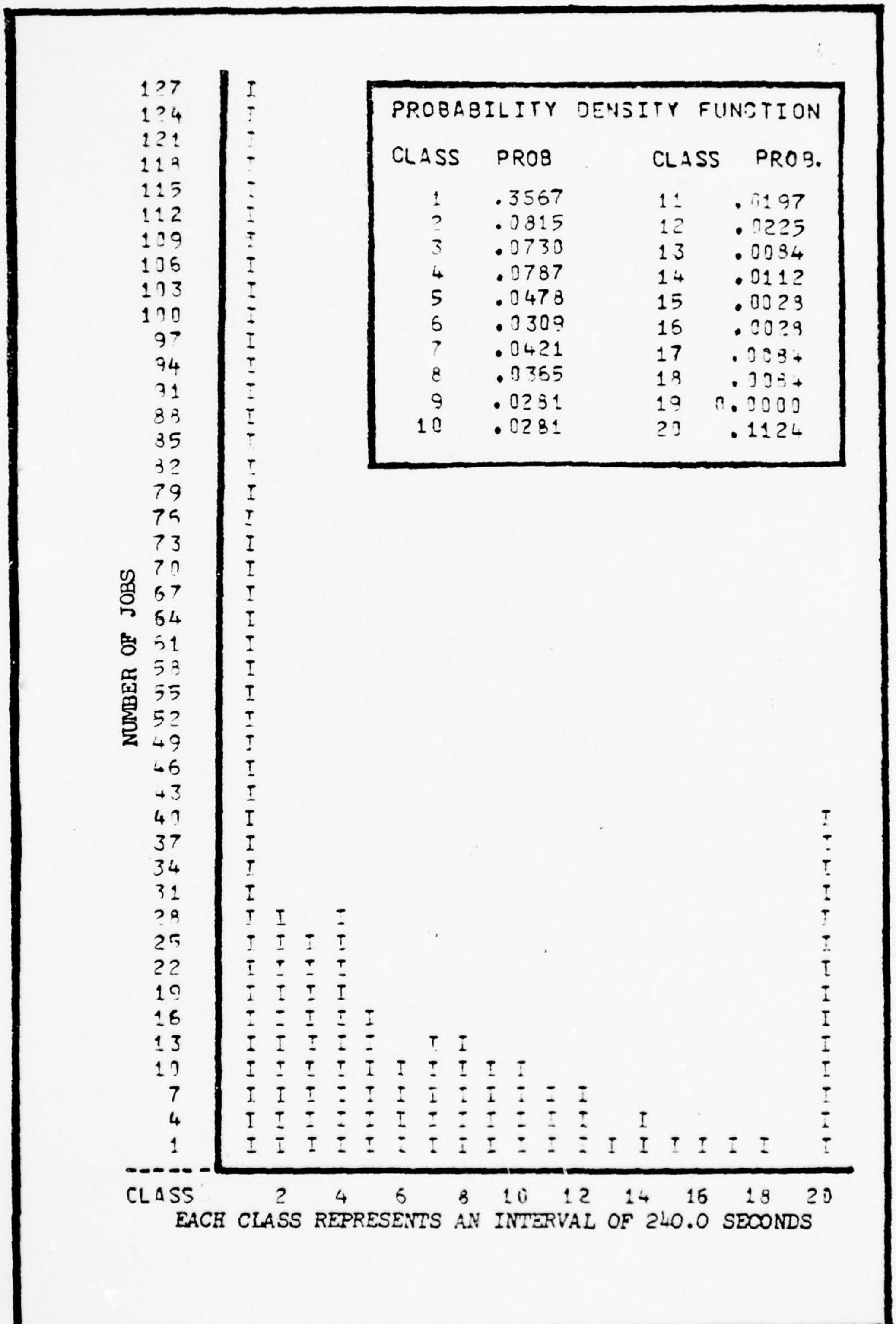
Fig. 31G. Distribution of Control Point Locations for 7 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.3455	11	.0112
2	.1067	12	.0169
3	.0730	13	.0225
4	.0702	14	.0225
5	.0590	15	.0169
6	.0534	16	0.0000
7	.0253	17	.0084
8	.0253	18	.0056
9	.0197	19	.0112
10	.0225	20	.0843

Fig. 31H. Distribution of Interarrival time for 7 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.3567	11	.0197
2	.0815	12	.0225
3	.0730	13	.0084
4	.0787	14	.0112
5	.0478	15	.0028
6	.0309	16	.0028
7	.0421	17	.0084
8	.0365	18	.0064
9	.0281	19	0.0000
10	.0281	20	.1124

Fig. 31I. Distribution of Input Queue time for 7 October

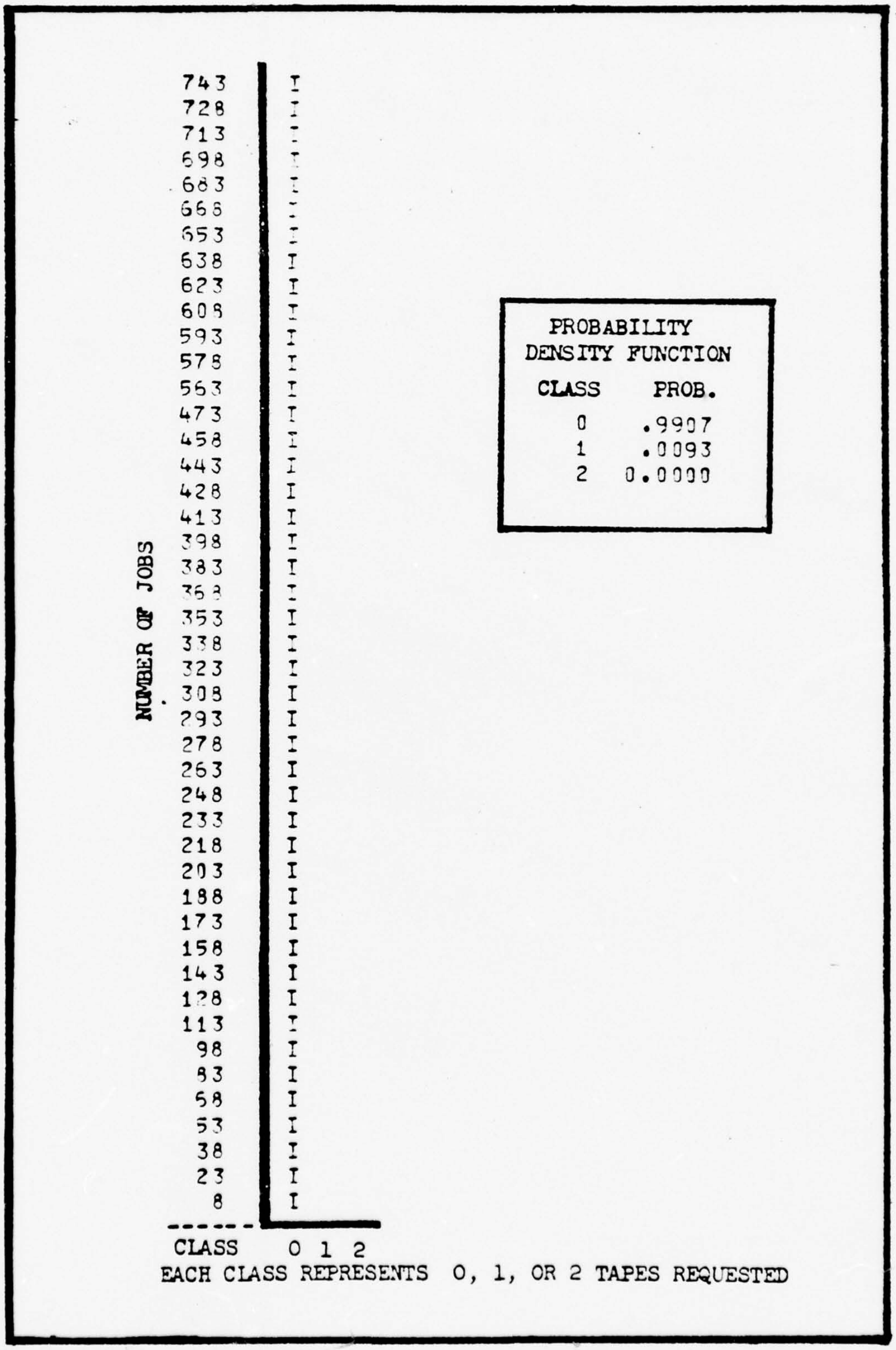


Fig. 31J. Distribution of Tapes Requested for 7 October

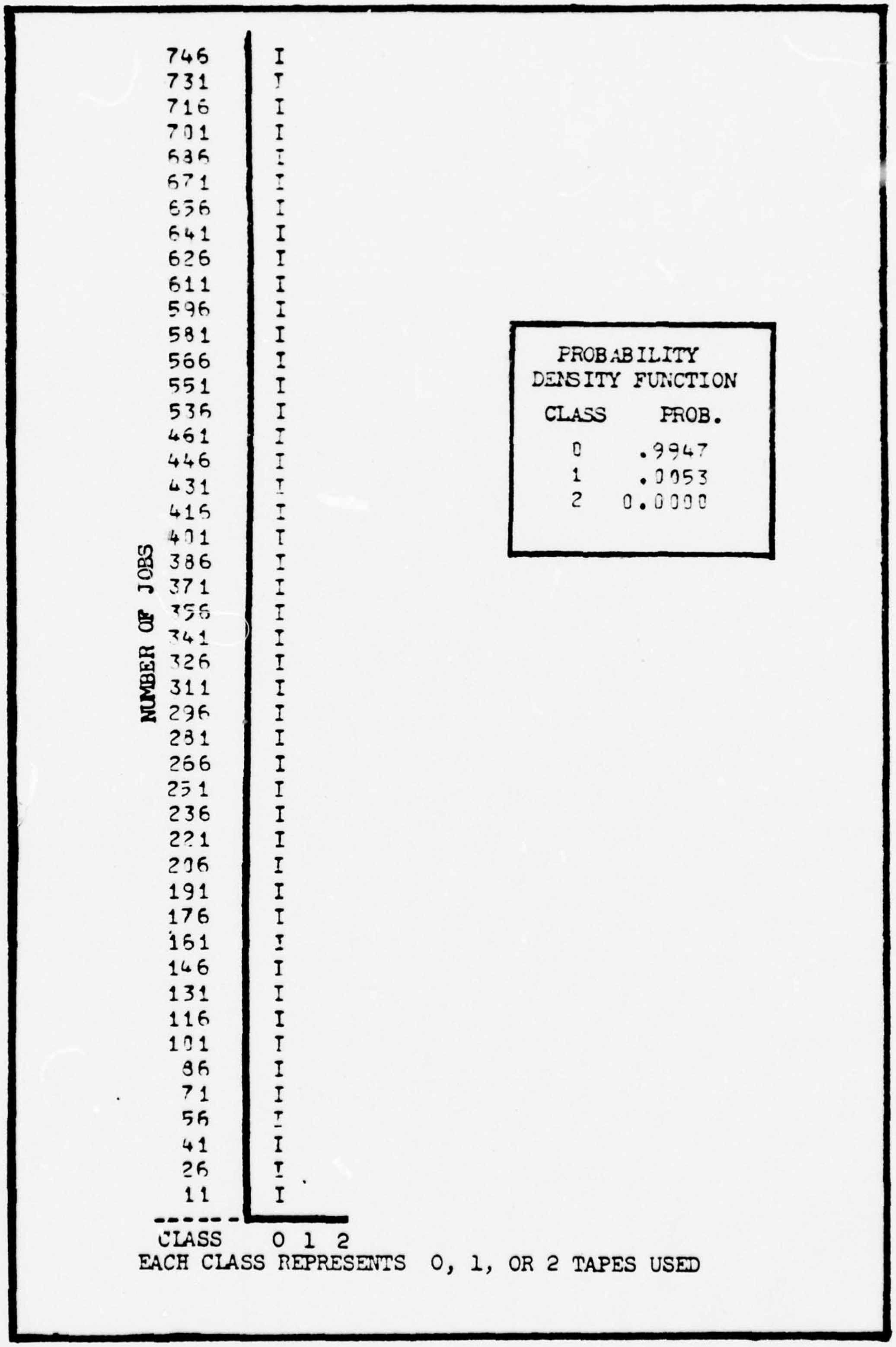


Fig. 31K. Distribution of Tapes Used for 7 October

TABLE XVI

Statistical Summary for Workload Parameters,

7 October

Variable CPTIME CPU time in seconds

Mean	23.309	Std Err	2.572	Std Dev	71.097
Variance	5054.795	Kurtosis	41.989	Skewness	5.622
Minimum	.003	Maximum	797.456	Sum	20864.204
C.V. Pct	260.341	.95 C.I.	22.260	to	32.359

Variable PPTIME Peripheral Processor time in seconds

Mean	98.032	Std Err	28.998	Std Dev	801.530
Variance	642450.031	Kurtosis	429.302	Skewness	19.234
Minimum	.312	Maximum	19197.203	Sum	74896.181
C.V. Pct	817.623	.95 C.I.	41.106	to	154.958

Variable TIMEIO Input-Output time in seconds

Mean	881.049	Std Err	127.519	Std Dev	3524.707
Variance	1242E+08	Kurtosis	179.228	Skewness	12.026
Minimum	.201	Maximum	60801.489	Sum	673121.623
C.V. Pct	400.056	.95 C.I.	630.719	to	1131.380

Variable TOTCOST Total cost in CRUs

Mean	31.461	Std Err	3.519	Std Dev	97.263
Variance	9460.083	Kurtosis	95.524	Skewness	8.572
Minimum	.043	Maximum	1398.792	Sum	24036.366
C.V. Pct	309.152	.95 C.I.	24.553	to	38.369

TABLE XVI (continued)

Variable KWS Memory in kilo-word seconds

Mean	881.252	Std Err	127.528	Std Dev	3524.958
Variance	1242E+08	Kurtosis	179.234	Skewness	12.026
Minimum	.167	Maximum	60806.511	Sum	673276.173
C.V. Pct	399.995	.95 C.I.	630.903	to	1131.600

Variable CPCT Control Point Occupancy time in seconds

Mean	804.031	Std Err	49.439	Std Dev	1366.526
Variance	1867393.121	Kurtosis	15.647	Skewness	3.353
Minimum	1.000	Maximum	12249.000	Sum	614280.000
C.V. Pct	169.959	.95 C.I.	706.976	to	901.084

Variable CMLOC Central Memory Locations in Use

Mean	149776.754	Std Err	2633.577	Std Dev	72793.505
Variance	5298E+10	Kurtosis	-.924	Skewness	-.064
Minimum	2048.000	Maximum	260096.000	Sum	.1144E+09
C.V. Pct	48.601	.95 C.I.	144606.838	to	154946.670

Variable CPLOC Control Points in Use

Mean	61.209	Std Err	1.002	Std Dev	27.687
Variance	766.564	Kurtosis	-1.375	Skewness	.185
Minimum	3.000	Maximum	100.000	Sum	46764.000
C.V. Pct	43.233	.95 C.I.	59.343	to	63.175

Variable ROLLOUT Total Time Rolled Out in seconds

Mean	22.463	Std Err	7.951	Std Dev	219.770
Variance	48299.001	Kurtosis	247.703	Skewness	14.605
Minimum	0.000	Maximum	4413.000	Sum	17162.000
C.V. Pct	978.351	.95 C.I.	6.855	to	38.072

TABLE XVI (continued)

Variable IATIME Inter-arrival time in seconds

Mean	65.345	Std Err	6.462	Std Dev	124.968
Variance	15616.956	Kurtosis	64.337	Skewness	6.430
Minimum	0.000	Maximum	1600.000	Sum	24439.000
C.V. Pct	191.243	.95 C.I.	52.639	to	78.051

Variable INQTIME Time in Input Queue in seconds

Mean	808.288	Std Err	74.217	Std Dev	1995.591
Variance	3982381.765	Kurtosis	35.014	Skewness	4.799
Minimum	0.000	Maximum	23975.000	Sum	584392.000
C.V. Pct	246.891	.95 C.I.	662.581	to	953.994

Variable TAPEREQ Number of Tapes Requested

Mean	.059	Std Err	.010	Std Dev	.272
Variance	.074	Kurtosis	26.764	Skewness	5.024
Minimum	0.000	Maximum	2.000	Sum	45.000
C.V. Pct	461.385	.95 C.I.	.040	to	.078

Variable TAPEDR Number of Tapes Used

Mean	.055	Std Err	.009	Std Dev	.250
Variance	.063	Kurtosis	25.341	Skewness	4.868
Minimum	0.000	Maximum	2.000	Sum	42.000
C.V. Pct	454.779	.95 C.I.	.037	to	.073

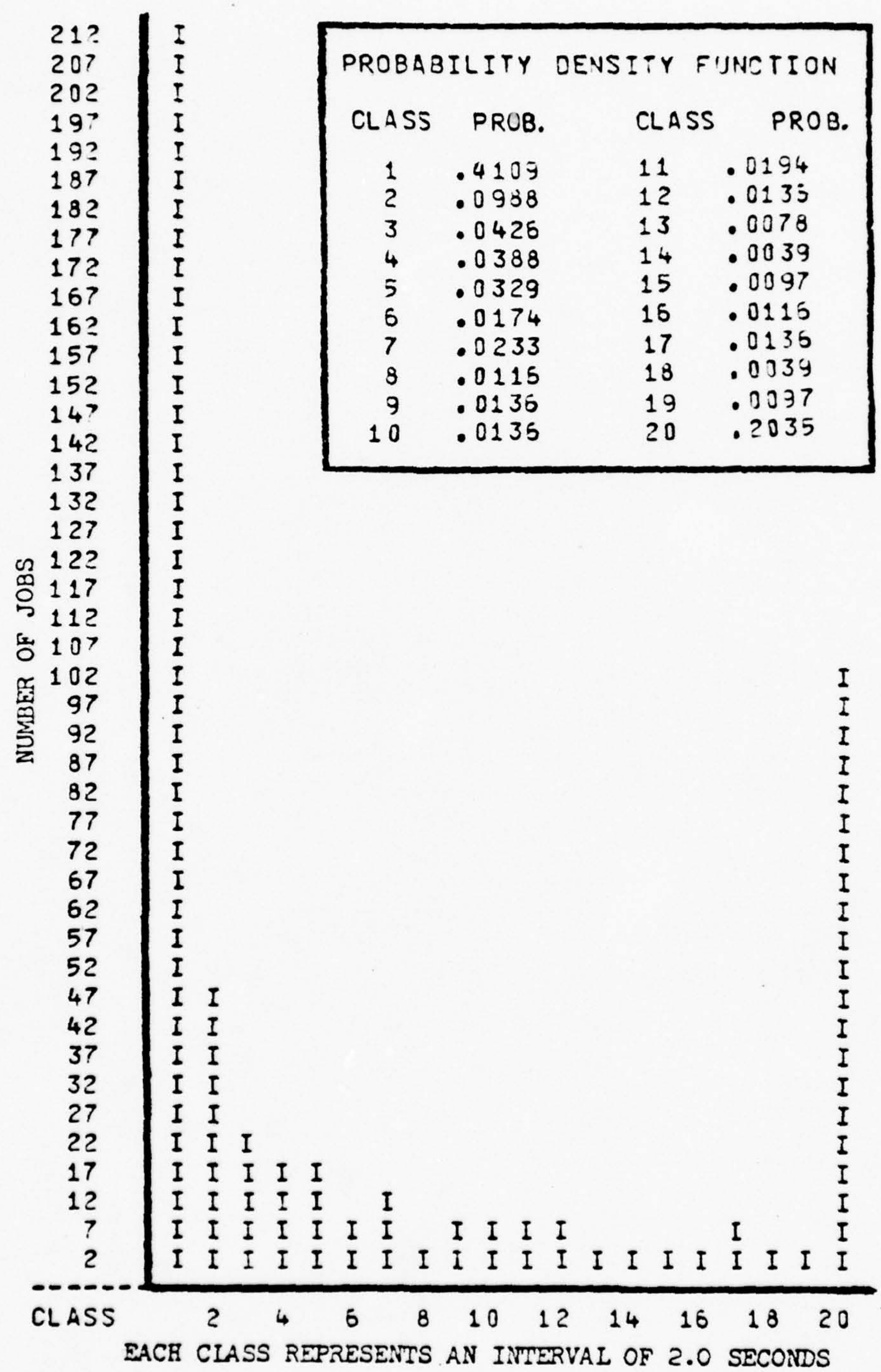
Appendix C

Workload Characterization,

29 September

The following variables are characterized with a frequency histogram, probability density function, and/or a statistical summary. All measurements of time are in seconds.

Name	Description
CPTIME	The central processor time needed to process a job.
PPTIME	The peripheral processor time needed to process a job.
TIMEIO	The total I/O time needed to process a job.
TOTCOST	The total job cost in terms of CRUs.
KSW	The memory usage in terms of kilo-word seconds.
CPOT	The total control point occupancy time which is the time a job enters a control point until it leaves a control point for the last time.
CMLOC	The number of central memory locations occupied by all jobs including this job at the time this job left a control point.
CPLOC	The number of control points occupied by all jobs including this job at the time this job left a control point.
ROLLOUT	The total time that a batch job was rolled out for processing magnetic tapes.
IATIME	The total time between this batch job and the last batch job to arrive into the input queue.
INQTIME	The time a batch job spends in the input queue.
TAPEREQ/TAPEUSED	The number of tapes requested/used per job.



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.4109	11	.0194
2	.0988	12	.0135
3	.0426	13	.0078
4	.0388	14	.0039
5	.0329	15	.0097
6	.0174	16	.0116
7	.0233	17	.0136
8	.0116	18	.0039
9	.0136	19	.0097
10	.0136	20	.2035

Fig 32A. Distribution of CPU time for 29 September

PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.1628	11	.0136
2	.2016	12	.0136
3	.0872	13	.0116
4	.0853	14	.0078
5	.0640	15	.0155
6	.0407	16	.0136
7	.0252	17	.0097
8	.0310	18	.0097
9	.0194	19	.0213
10	.0213	20	.1453

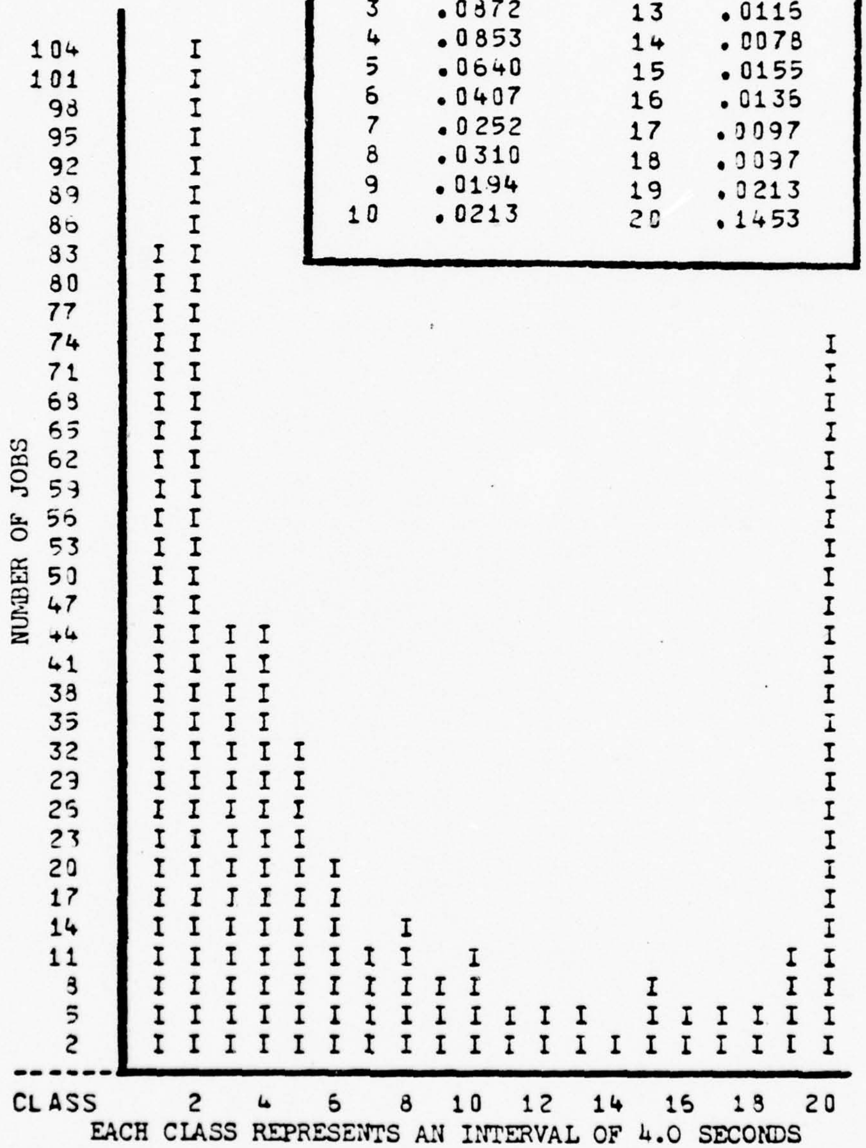
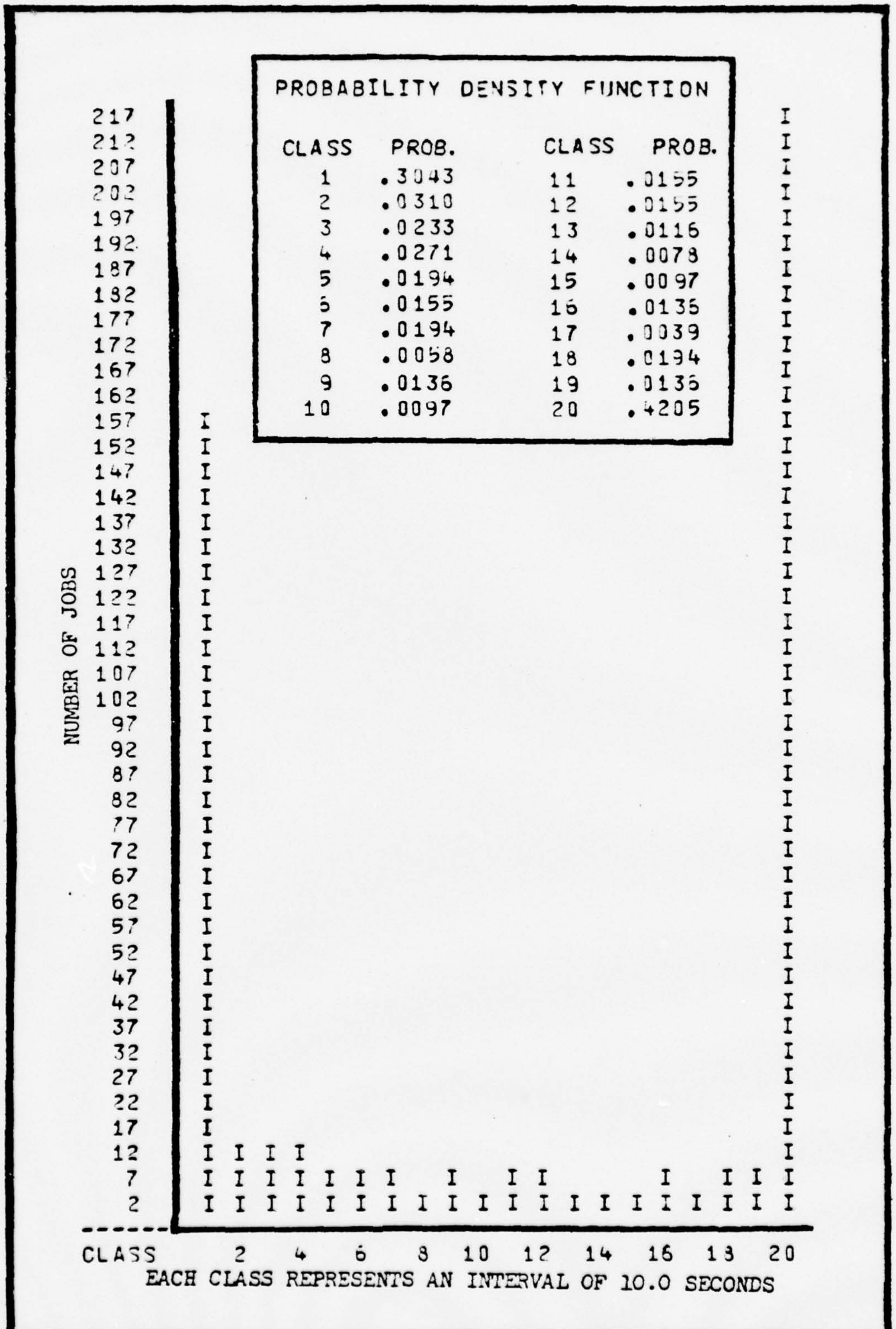


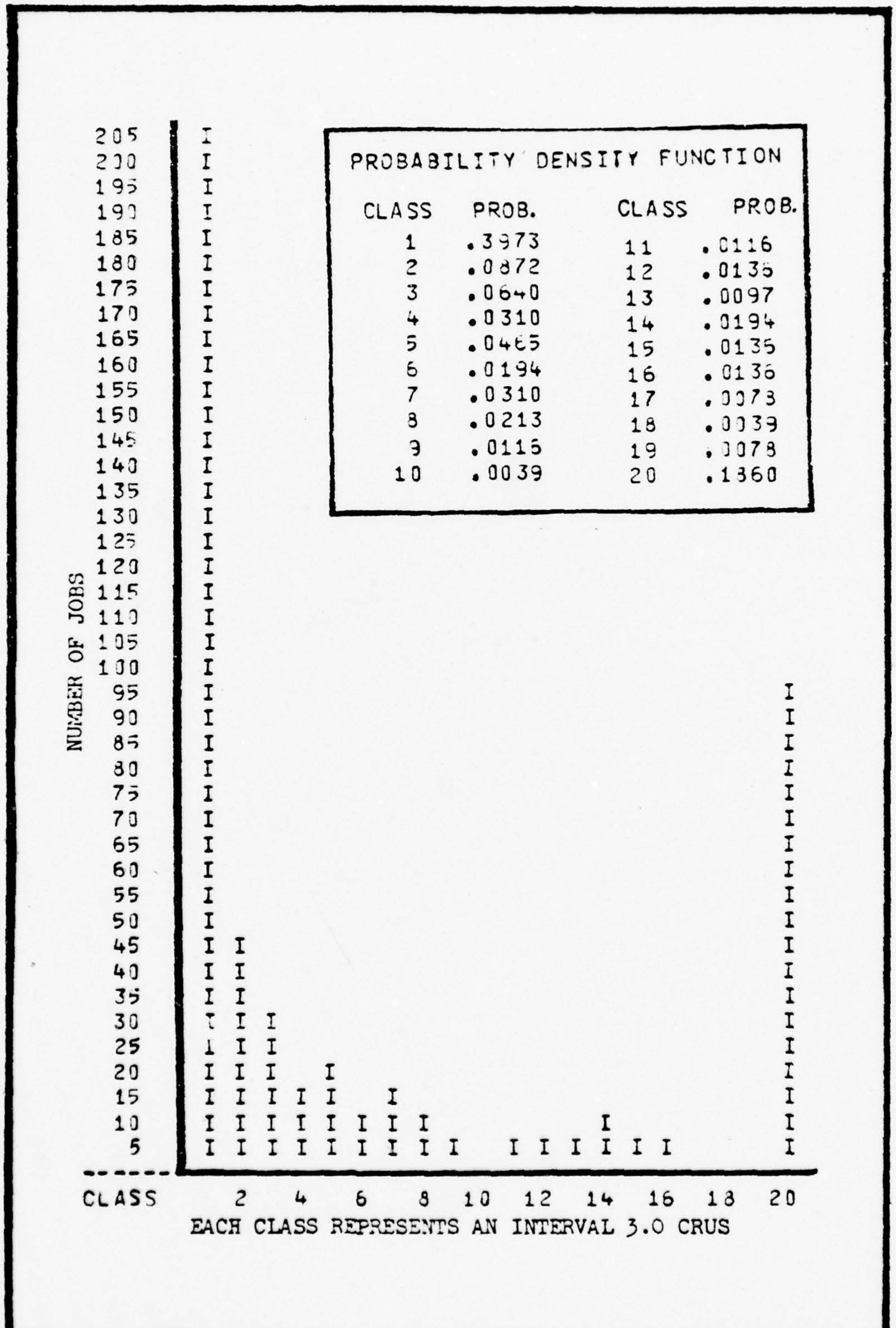
Fig. 32B. Distribution of PPU time for 29 September



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.3043	11	.0155
2	.0310	12	.0155
3	.0233	13	.0116
4	.0271	14	.0078
5	.0194	15	.0097
6	.0155	16	.0136
7	.0194	17	.0039
8	.0058	18	.0194
9	.0136	19	.0136
10	.0097	20	.4205

Fig. 32C. Distribution of I/O time for 29 September



CLASS	PROB.	CLASS	PROB.
1	.3973	11	.0116
2	.0872	12	.0136
3	.0640	13	.0097
4	.0310	14	.0194
5	.0465	15	.0135
6	.0194	16	.0136
7	.0310	17	.0078
8	.0213	18	.0039
9	.0116	19	.0078
10	.0039	20	.1860

Fig. 32D. Distribution of CRUs for 29 September

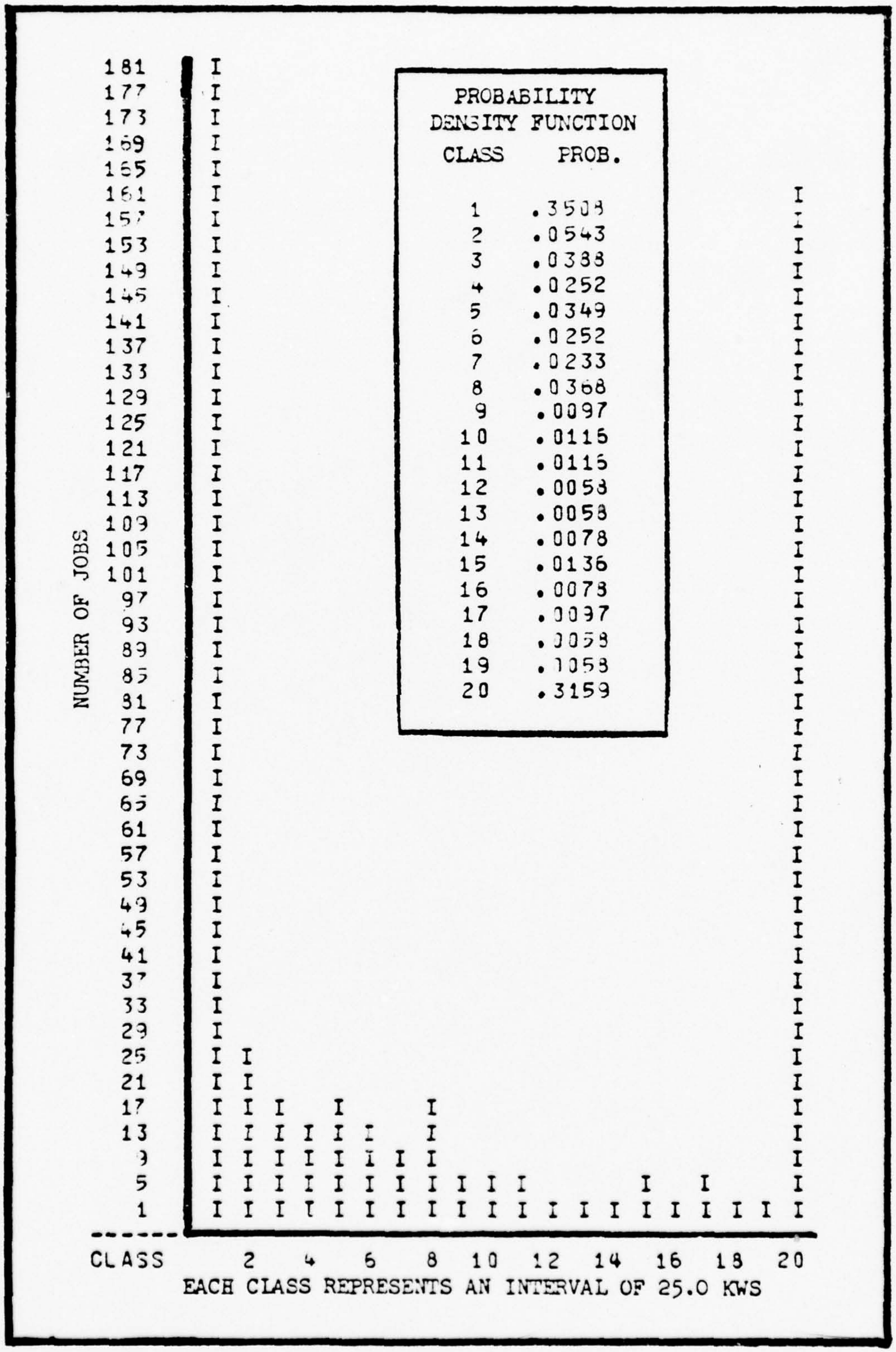


Fig. 32E. Distribution of KWS for 29 September

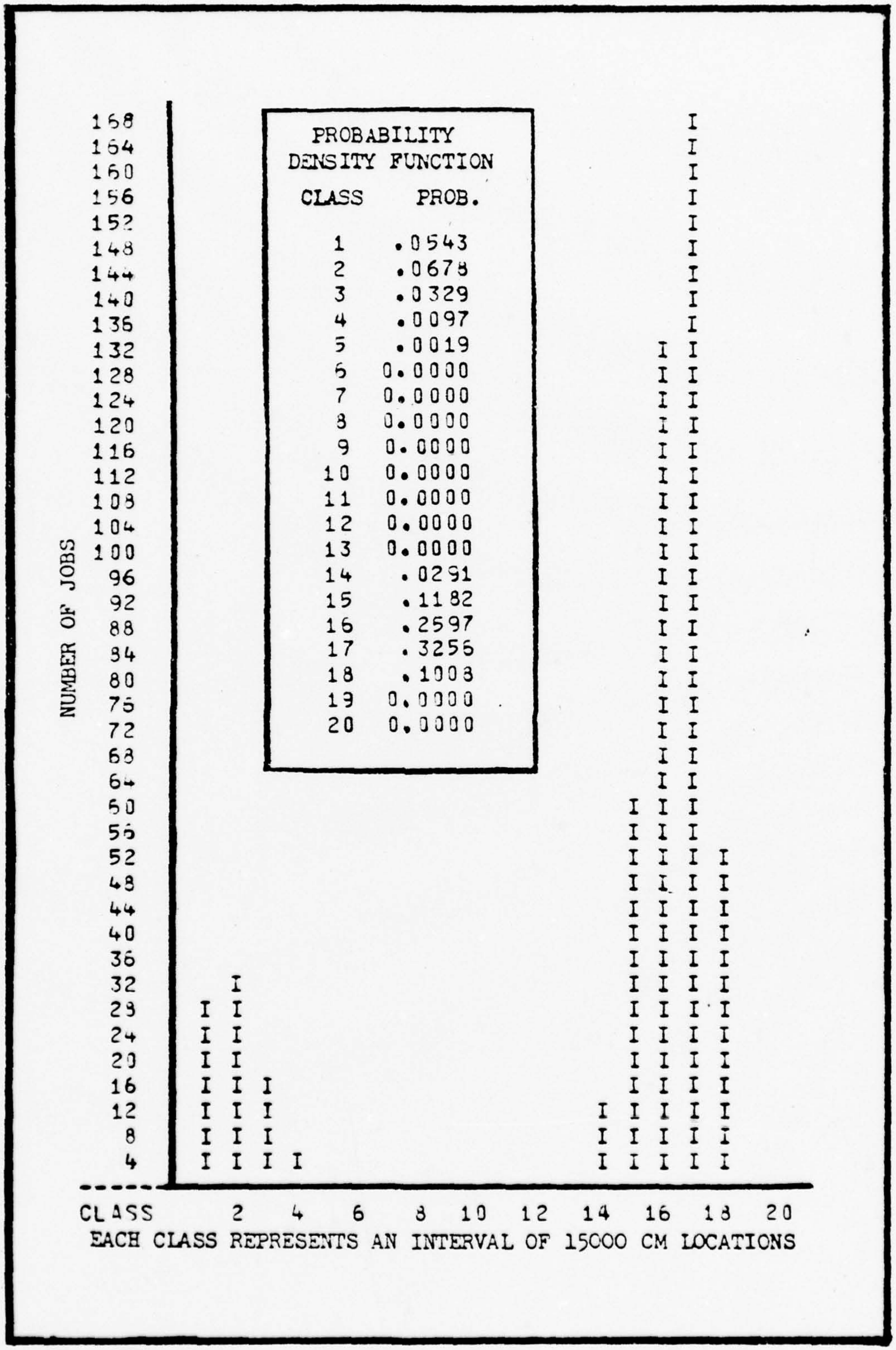


Fig. 32F. Distribution of Central Memory Locations for 29 September

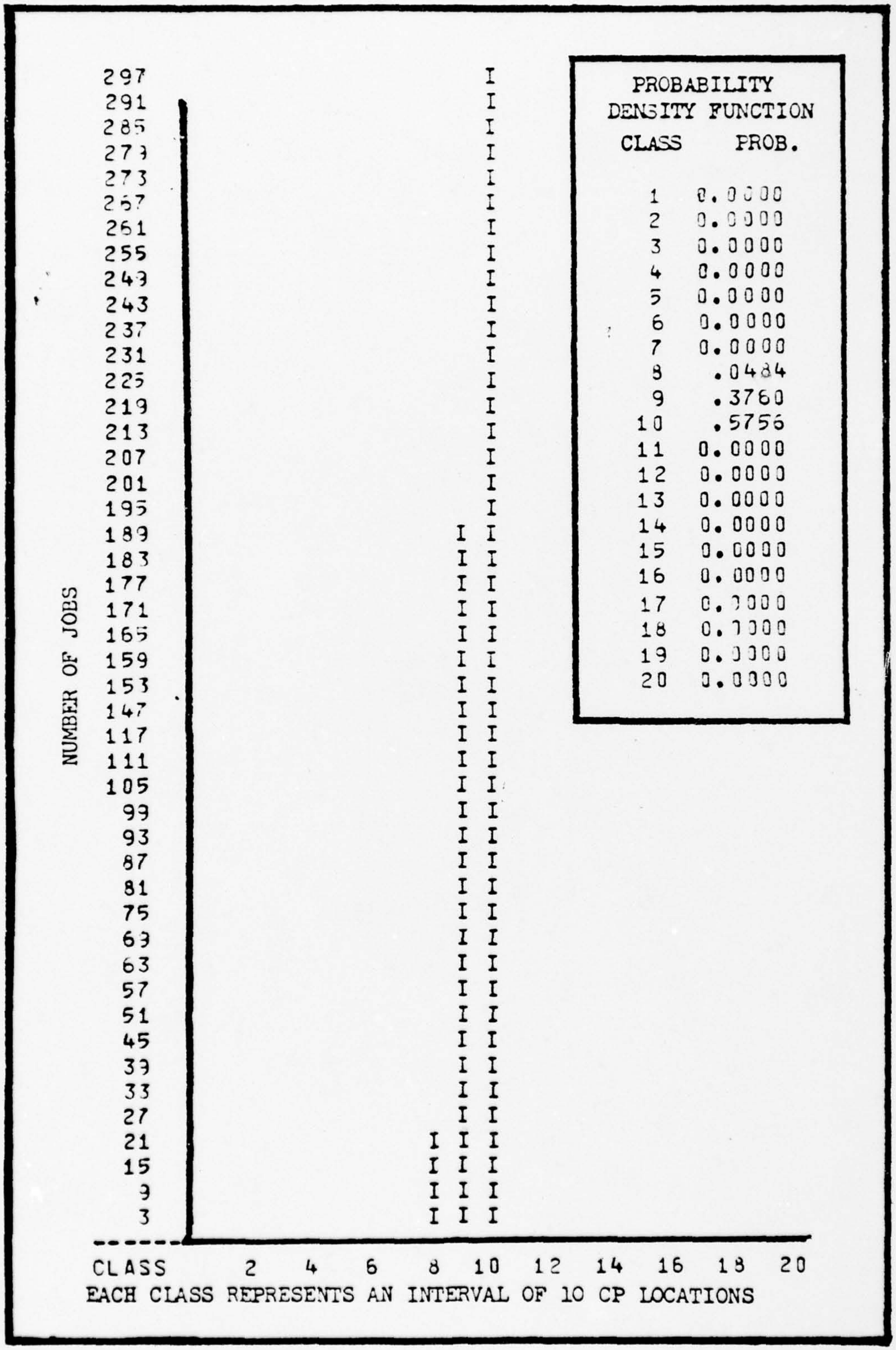
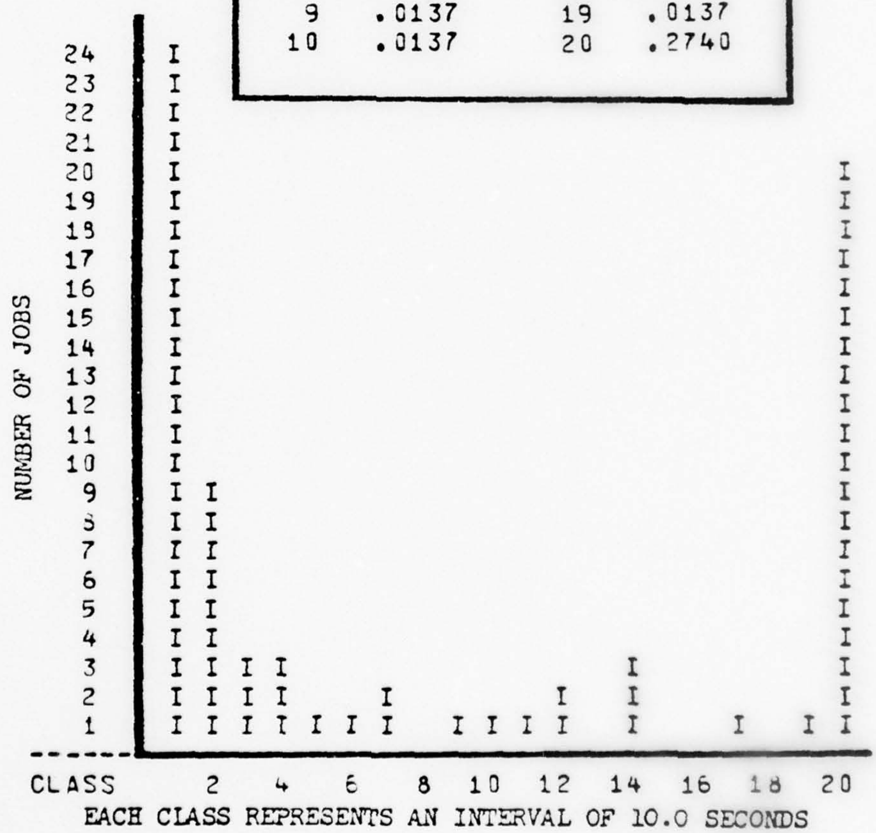


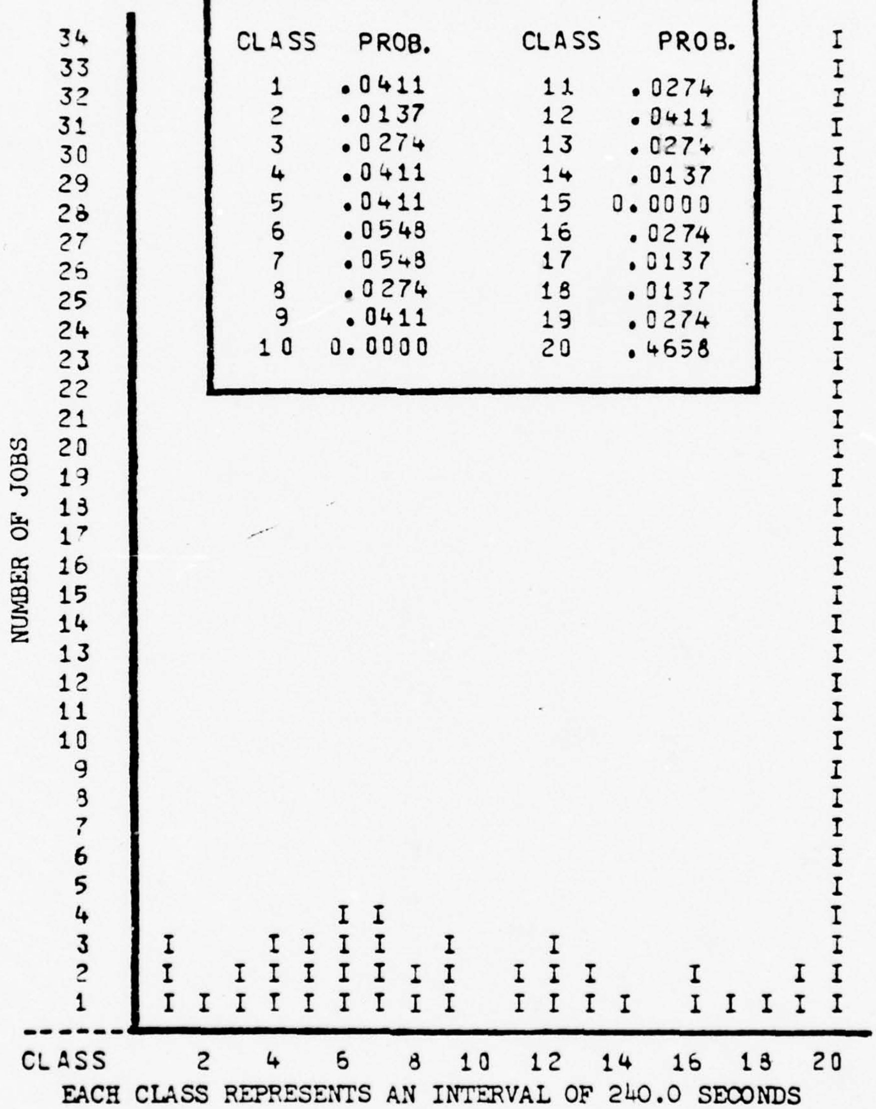
Fig. 32G. Distribution of Control Point Locations for 29 September



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.3288	11	.0137
2	.1233	12	.0274
3	.0411	13	0.0000
4	.0411	14	.0411
5	.0137	15	0.0000
6	.0137	16	0.0000
7	.0274	17	.0137
8	0.0000	18	0.0000
9	.0137	19	.0137
10	.0137	20	.2740

Fig. 32H. Distribution of Interarrival time for 29 September



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.0411	11	.0274
2	.0137	12	.0411
3	.0274	13	.0274
4	.0411	14	.0137
5	.0411	15	0.0000
6	.0548	16	.0274
7	.0548	17	.0137
8	.0274	18	.0137
9	.0411	19	.0274
10	0.0000	20	.4658

Fig. 32I. Distribution of Input Queue time for 29 September

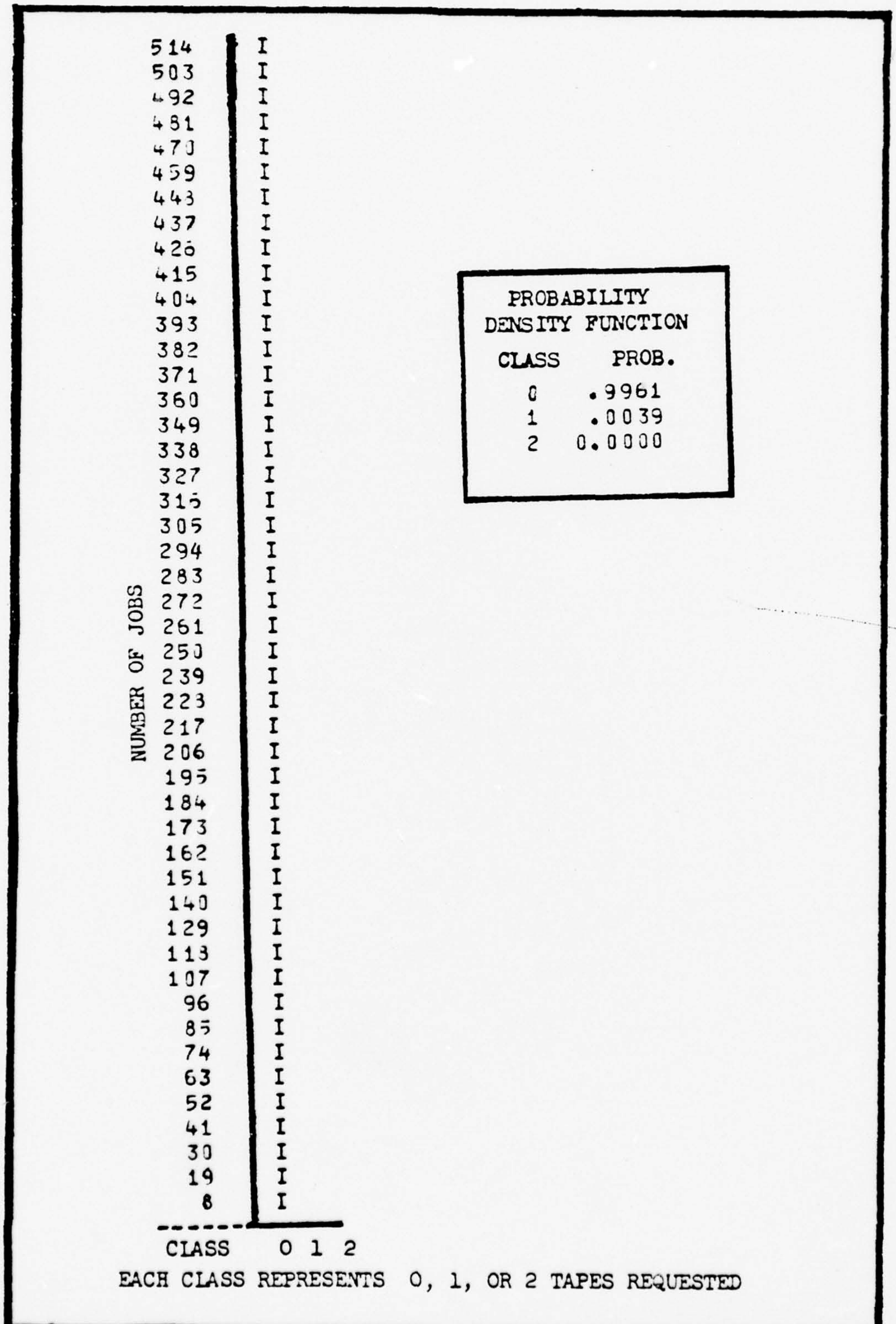


Fig. 32J. Distribution of Tapes Requested for 29 September

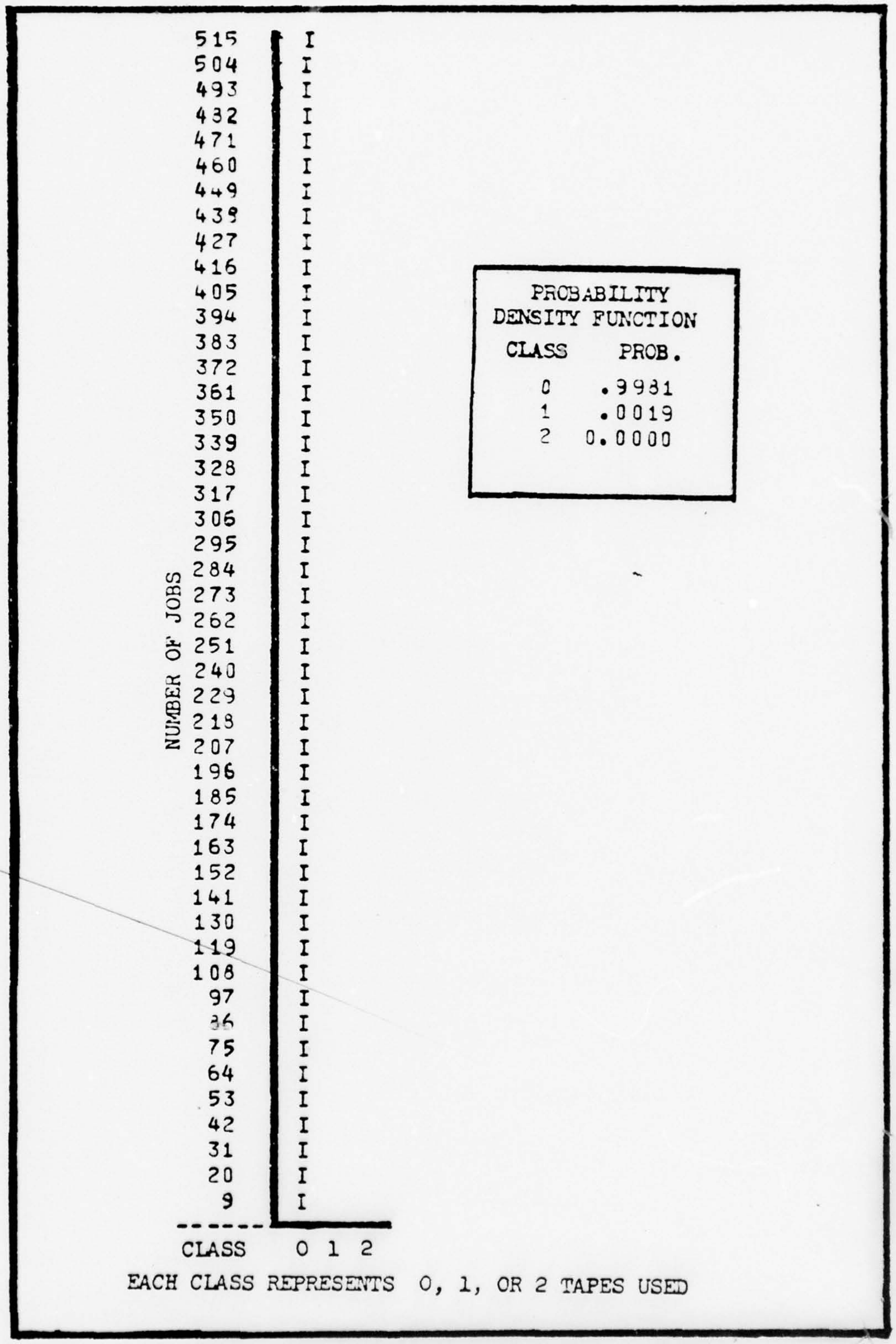


Fig. 32K. Distribution of Tapes Used for 29 September

TABLE XVII

Statistical Summary for Workload Parameters,

29 September

Variable CPTIME CPU time in seconds

Mean	35.165	Std Err	4.110	Std Dev	94.179
Variance	8869.739	Kurtosis	84.196	Skewness	7.681
Minimum	.008	Maximum	1329.140	Sum	18461.562
C.V. Pct	267.822	.95 C.I.	27.090	to	43.240

Variable PPTIME Peripheral Processor time in seconds

Mean	95.724	Std Err	30.235	Std Dev	692.773
Variance	479935.004	Kurtosis	191.164	Skewness	13.608
Minimum	1.629	Maximum	10649.088	Sum	50254.868
C.V. Pct	723.723	.95 C.I.	36.327	to	155.120

Variable TIMEIO Input-Output time in seconds

Mean	1397.715	Std Err	267.525	Std Dev	6129.771
Variance	3757E+08	Kurtosis	69.809	Skewness	8.153
Minimum	.237	Maximum	65373.390	Sum	733800.345
C.V. Pct	438.557	.95 C.I.	872.161	to	1923.268

Variable TOTCOST Total Cost in CRUs

Mean	43.550	Std Err	5.880	Std Dev	134.739
Variance	18154.380	Kurtosis	53.045	Skewness	6.919
Minimum	.125	Maximum	1311.826	Sum	22863.882
C.V. Pct	309.385	.95 C.I.	31.998	to	55.102

TABLE XVII (continued)

Variable KWS Memory in kilo-word seconds

Mean	1397.780	Std Err	267.547	Std Dev	6130.282
Variance	3758E+08	Kurtosis	69.809	Skewness	8.153
Minimum	.167	Maximum	65378.819	Sum	733834.565
C.V. Pct	438.573	.95 C.I.	872.183	to	1923.373

Variable CPCT Control Point Occupancy time in seconds

Mean	1158.486	Std Err	109.206	Std Dev	2502.221
Variance	6261109.048	Kurtosis	23.027	Skewness	4.172
Minimum	1.000	Maximum	24171.000	Sum	608205.000
C.V. Pct	215.991	.95 C.I.	943.951	to	1373.021

Variable CMLOC Central Memory Locations in Use

Mean	202637.897	Std Err	3551.593	Std Dev	81377.221
Variance	.6622E+10	Kurtosis	1.202	Skewness	-1.734
Minimum	512.000	Maximum	261632.000	Sum	.1063E+09
C.V. Pct	40.159	.95 C.I.	195650.787	to	209615.007

Variable CPLOC Control Points in Use

Mean	92.189	Std Err	.262	Std Dev	6.000
Variance	36.104	Kurtosis	-.777	Skewness	-.401
Minimum	79.000	Maximum	100.000	Sum	48399.000
C.V. Pct	6.518	.95 C.I.	91.673	to	92.704

Variable ROLLOUT Total Time Rolled Out in seconds

Mean	61.286	Std Err	23.785	Std Dev	544.939
Variance	297013.441	Kurtosis	188.328	Skewness	13.101
Minimum	0.000	Maximum	8944.000	Sum	32175.000
C.V. Pct	889.260	.95 C.I.	14.559	to	108.012

TABLE XVII (continued)

Variable IATIME Inter-arrival time in seconds

Mean	49.918	Std Err	11.415	Std Dev	195.067
Variance	38050.969	Kurtosis	34.897	Skewness	5.565
Minimum	0.000	Maximum	1607.000	Sum	14567.000
C.V. Pct	390.776	.95 C.I.	27.451	to	72.385

Variable INQTIME Time in Input Queue in seconds

Mean	799.853	Std Err	124.053	Std Dev	2723.521
Variance	7417565.756	Kurtosis	26.617	Skewness	4.800
Minimum	0.000	Maximum	20661.000	Sum	385529.000
C.V. Pct	740.503	.95 C.I.	556.100	to	1043.605

Variable TAPEREQ Number of Tapes Requested

Mean	.078	Std Err	.012	Std Dev	.282
Variance	.080	Kurtosis	13.237	Skewness	3.642
Minimum	0.000	Maximum	2.000	Sum	41.000
C.V. Pct	361.650	.95 C.I.	.054	to	.102

Variable TAPEUSED Number of Tapes Used

Mean	.076	Std Err	.012	Std Dev	.273
Variance	.074	Kurtosis	11.332	Skewness	3.474
Minimum	0.000	Maximum	2.000	Sum	40.000
C.V. Pct	357.850	.95 C.I.	.053	to	.100

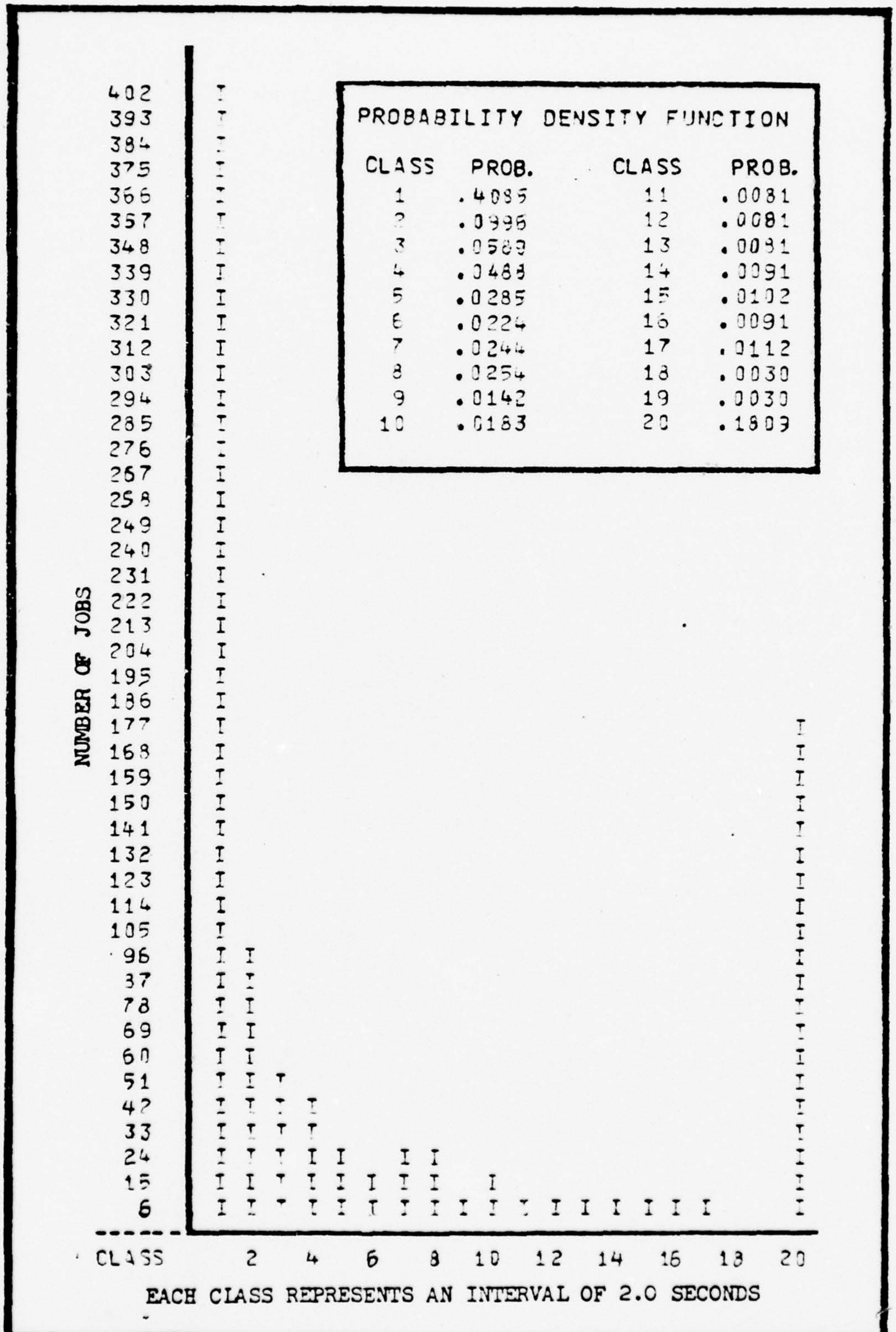
Appendix D

Workload Characterization

25 October

The following variables are characterized with a frequency histogram, probability density function, and/or a statistical summary. All measurements of time are in seconds.

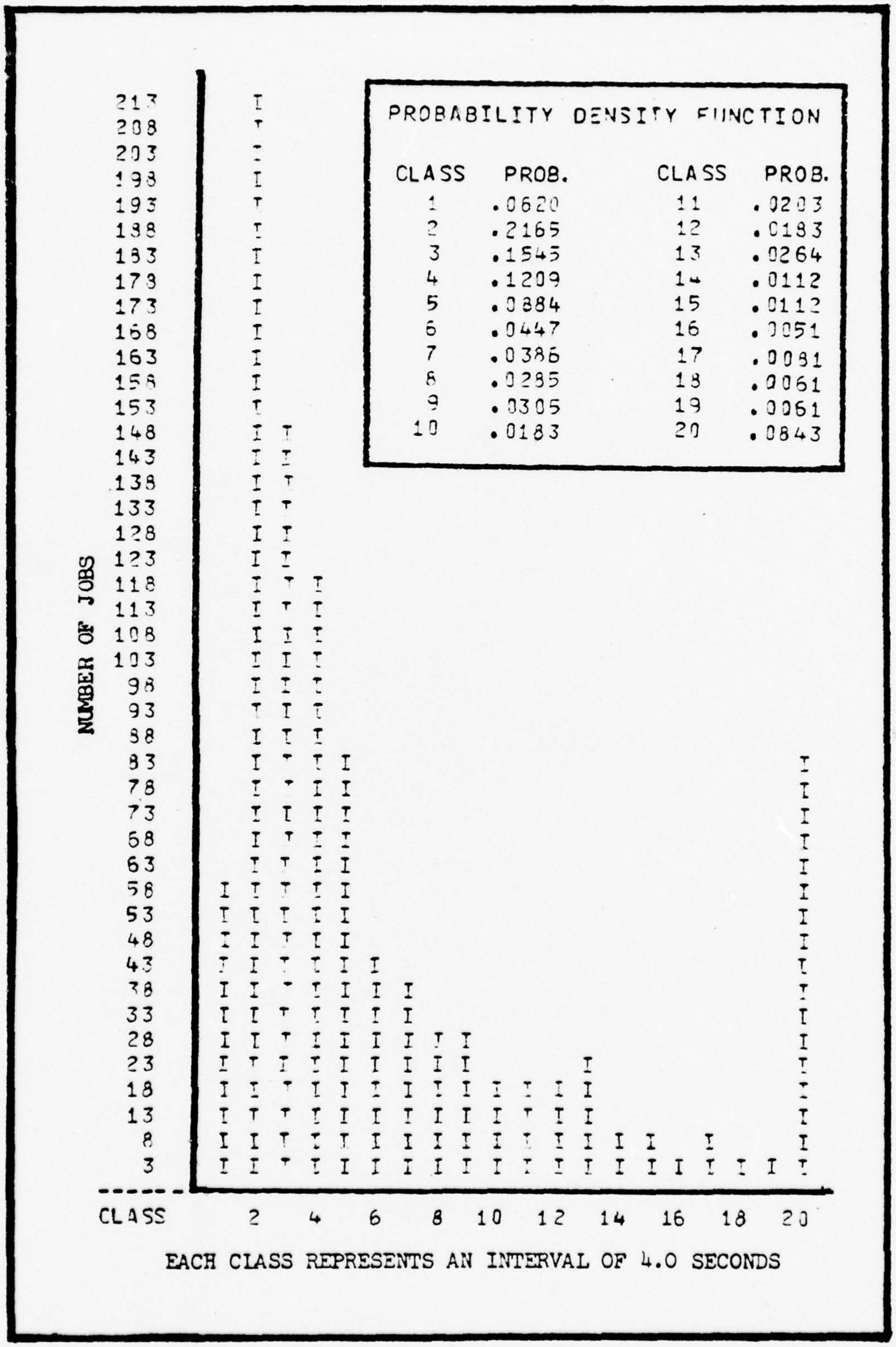
Name	Description
CPTIME	The central processor time needed to process a job.
PPTIME	The peripheral processor time needed to process a job.
TIMEIO	The total I/O time needed to process a job.
TOTCOST	The total job cost in terms of CRUs.
KSW	The memory usage in terms of kilo-word seconds.
CPCT	The total control point occupancy time which is the time a job enters a control point until it leaves a control point for the last time.
CMLOC	The number of central memory locations occupied by all jobs including this job at the time this job left a control point.
CPLOC	The number of control points occupied by all jobs including this job at the time this job left a control point.
ROLLOUT	The total time that a batch job was rolled out for processing magnetic tapes.
IATIME	The total time between this batch job and the last batch job to arrive into the input queue.
INQTIME	The time a batch job spends in the input queue.
TAPEREQ/TAPEUSED	The number of tapes requested/used per job.



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.4085	11	.0081
2	.0996	12	.0081
3	.0589	13	.0081
4	.0488	14	.0091
5	.0285	15	.0102
6	.0224	16	.0091
7	.0244	17	.0112
8	.0254	18	.0030
9	.0142	19	.0030
10	.0183	20	.1809

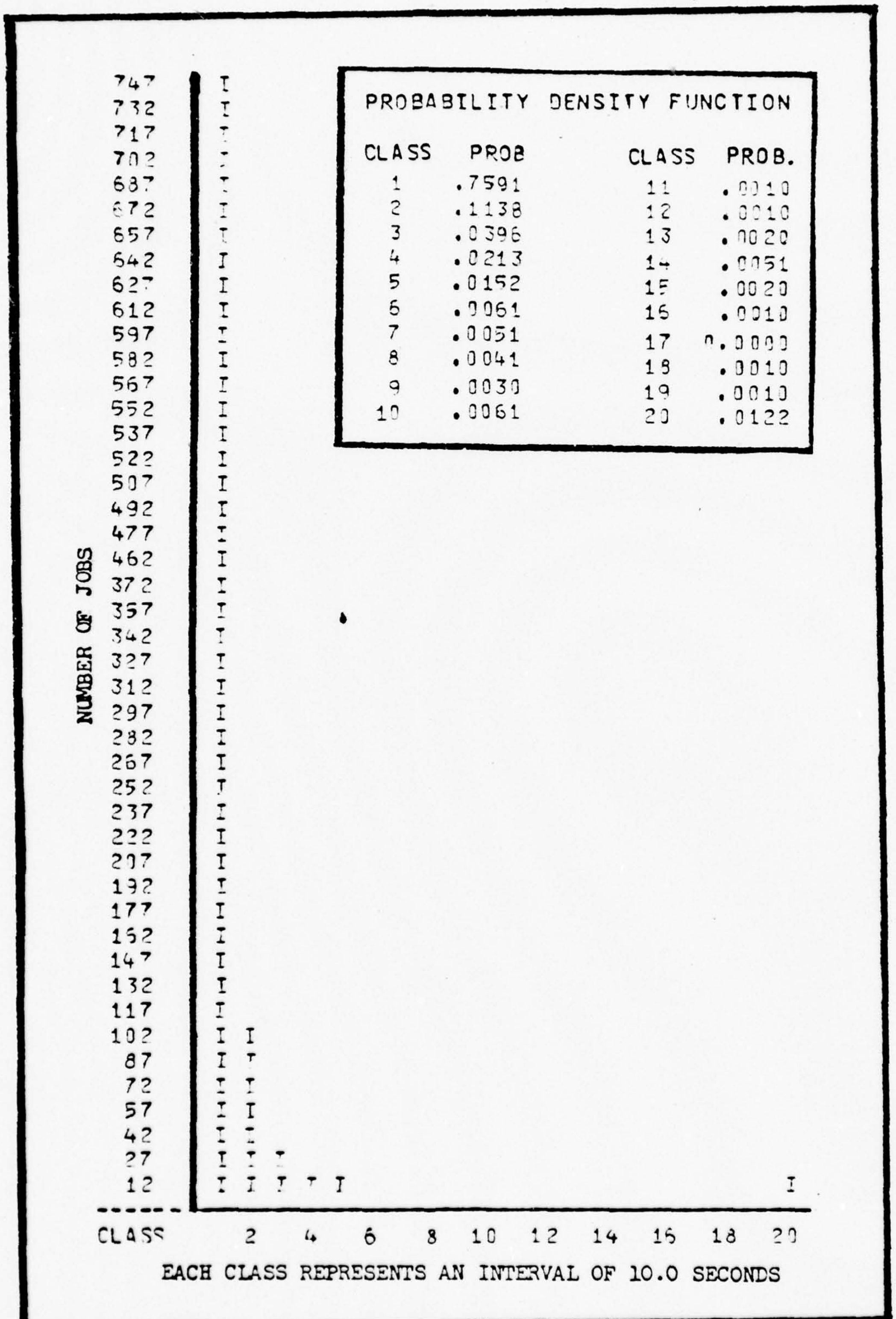
Fig. 33A. Distribution of CPU time for 25 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.0620	11	.0203
2	.2165	12	.0183
3	.1545	13	.0264
4	.1209	14	.0112
5	.0884	15	.0112
6	.0447	16	.0051
7	.0386	17	.0081
8	.0285	18	.0061
9	.0305	19	.0061
10	.0183	20	.0843

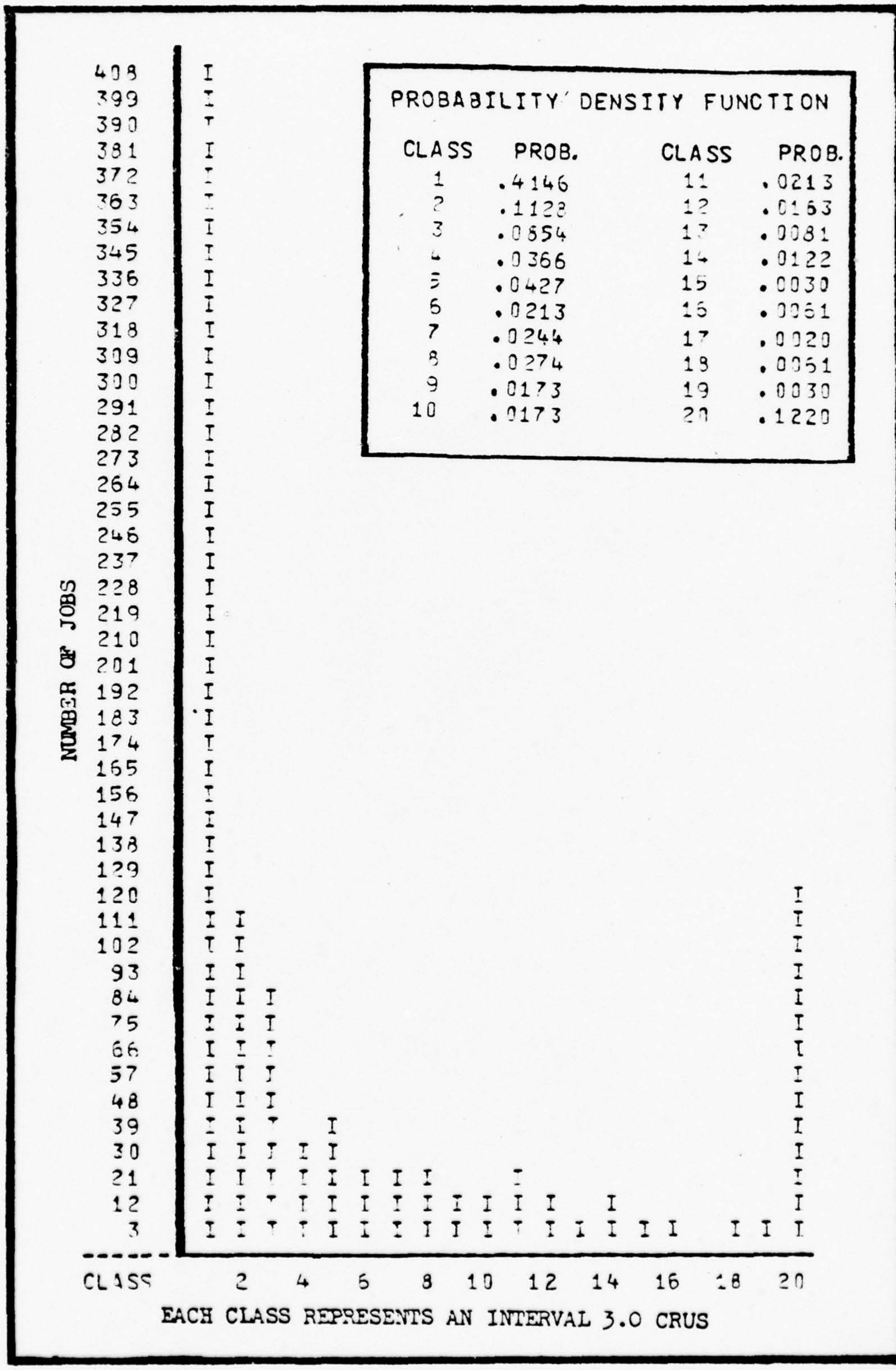
Fig. 33B. Distribution of PPU time for 25 October



PROBABILITY DENSITY FUNCTION

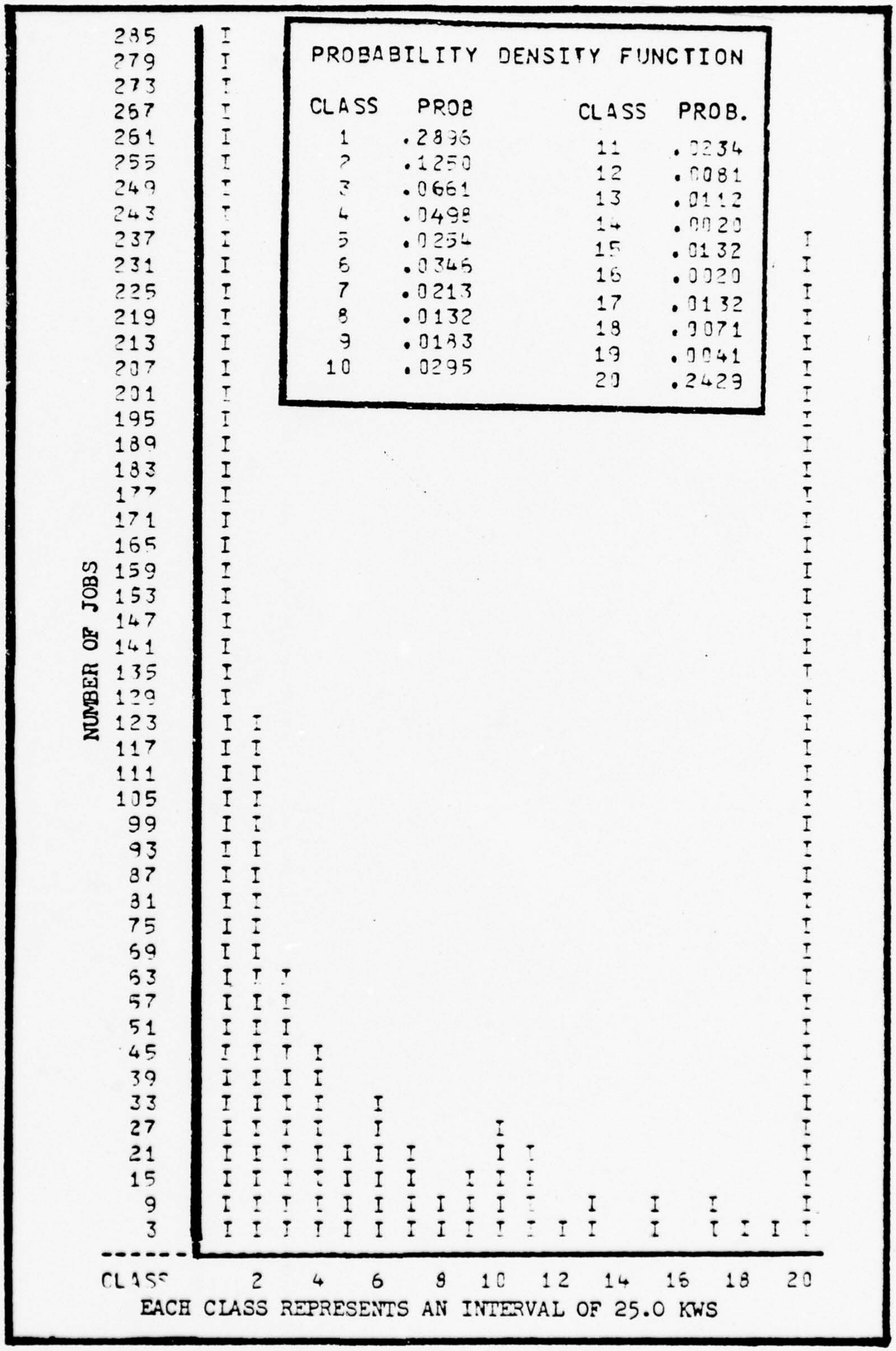
CLASS	PROB	CLASS	PROB.
1	.7591	11	.0010
2	.1138	12	.0010
3	.0396	13	.0020
4	.0213	14	.0051
5	.0152	15	.0020
6	.0061	16	.0010
7	.0051	17	0.0000
8	.0041	18	.0010
9	.0030	19	.0010
10	.0061	20	.0122

Fig. 33C. Distribution of I/O time for 25 October



CLASS	PROB.	CLASS	PROB.
1	.4146	11	.0213
2	.1128	12	.0163
3	.0854	13	.0081
4	.0366	14	.0122
5	.0427	15	.0030
6	.0213	16	.0061
7	.0244	17	.0020
8	.0274	18	.0061
9	.0173	19	.0030
10	.0173	20	.1220

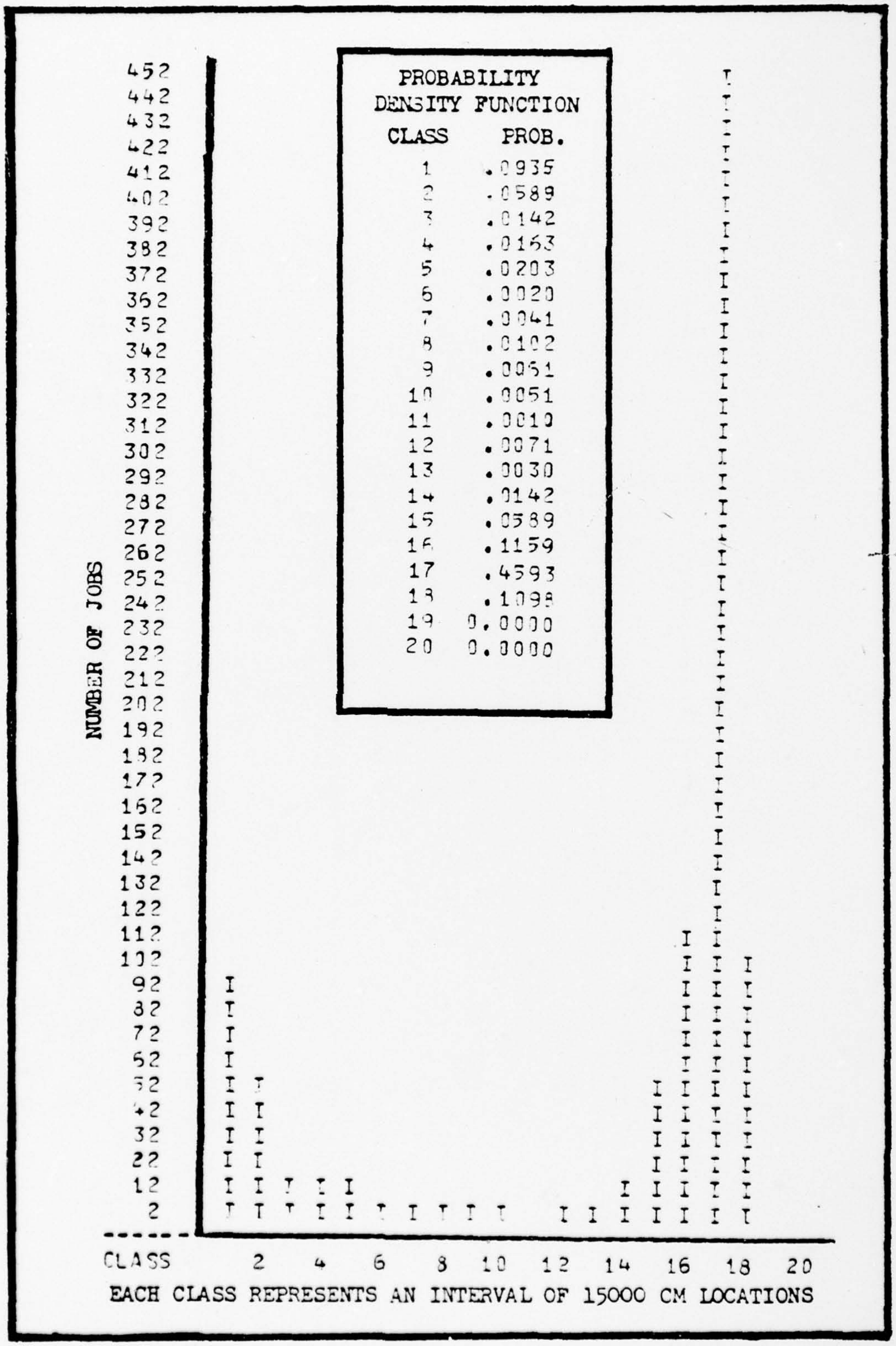
Fig. 33D. Distribution of CRUs for 25 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.2896	11	.0234
2	.1250	12	.0081
3	.0661	13	.0112
4	.0498	14	.0020
5	.0254	15	.0132
6	.0346	16	.0020
7	.0213	17	.0132
8	.0132	18	.0071
9	.0183	19	.0041
10	.0295	20	.2429

Fig. 33E. Distribution of KWS for 25 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.
1	.0935
2	.0589
3	.0142
4	.0163
5	.0203
6	.0020
7	.0041
8	.0102
9	.0061
10	.0051
11	.0010
12	.0071
13	.0030
14	.0142
15	.0589
16	.1159
17	.4593
18	.1098
19	0.0000
20	0.0000

Fig. 33F. Distribution of Central Memory Locations for 25 October

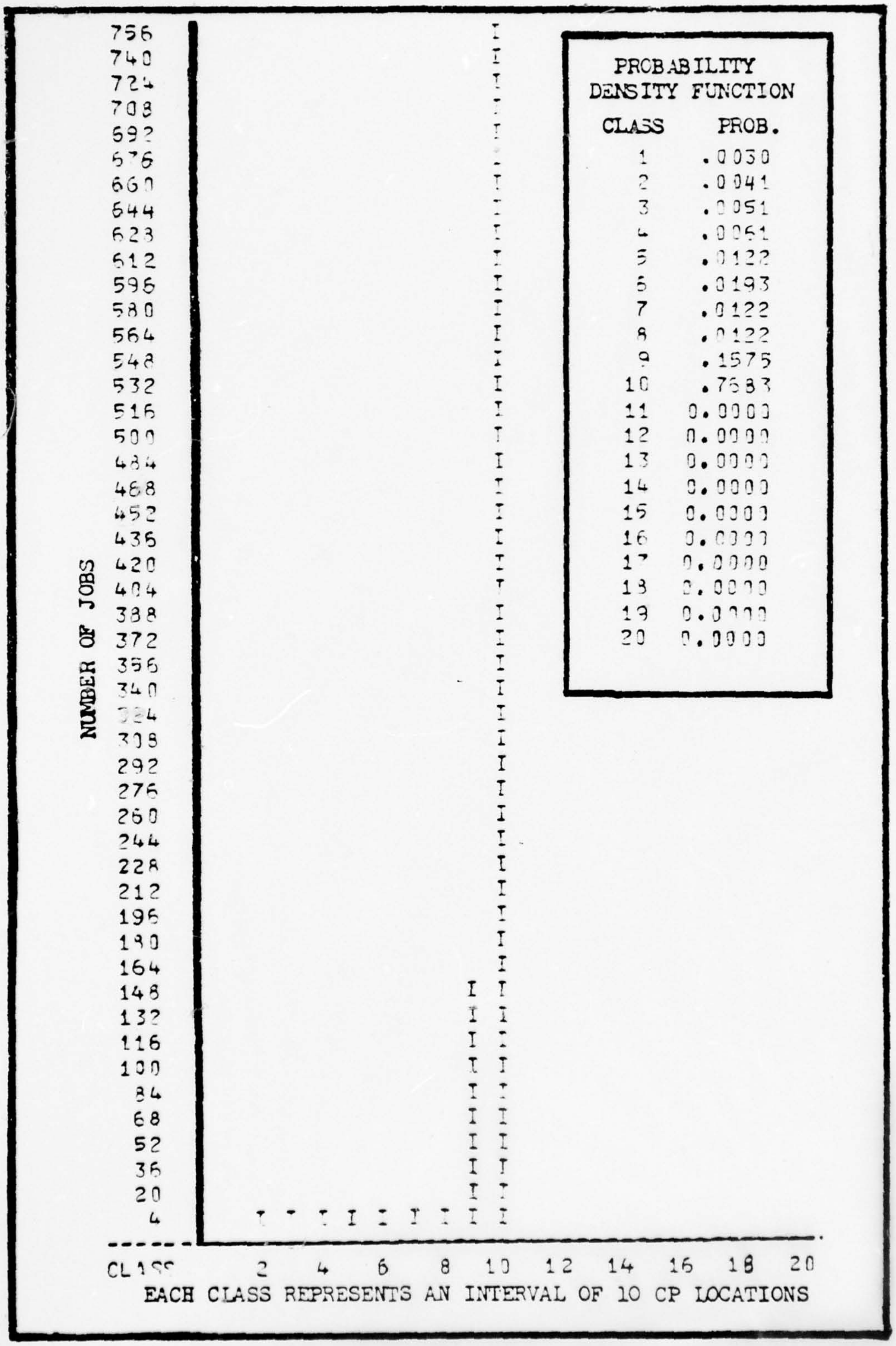
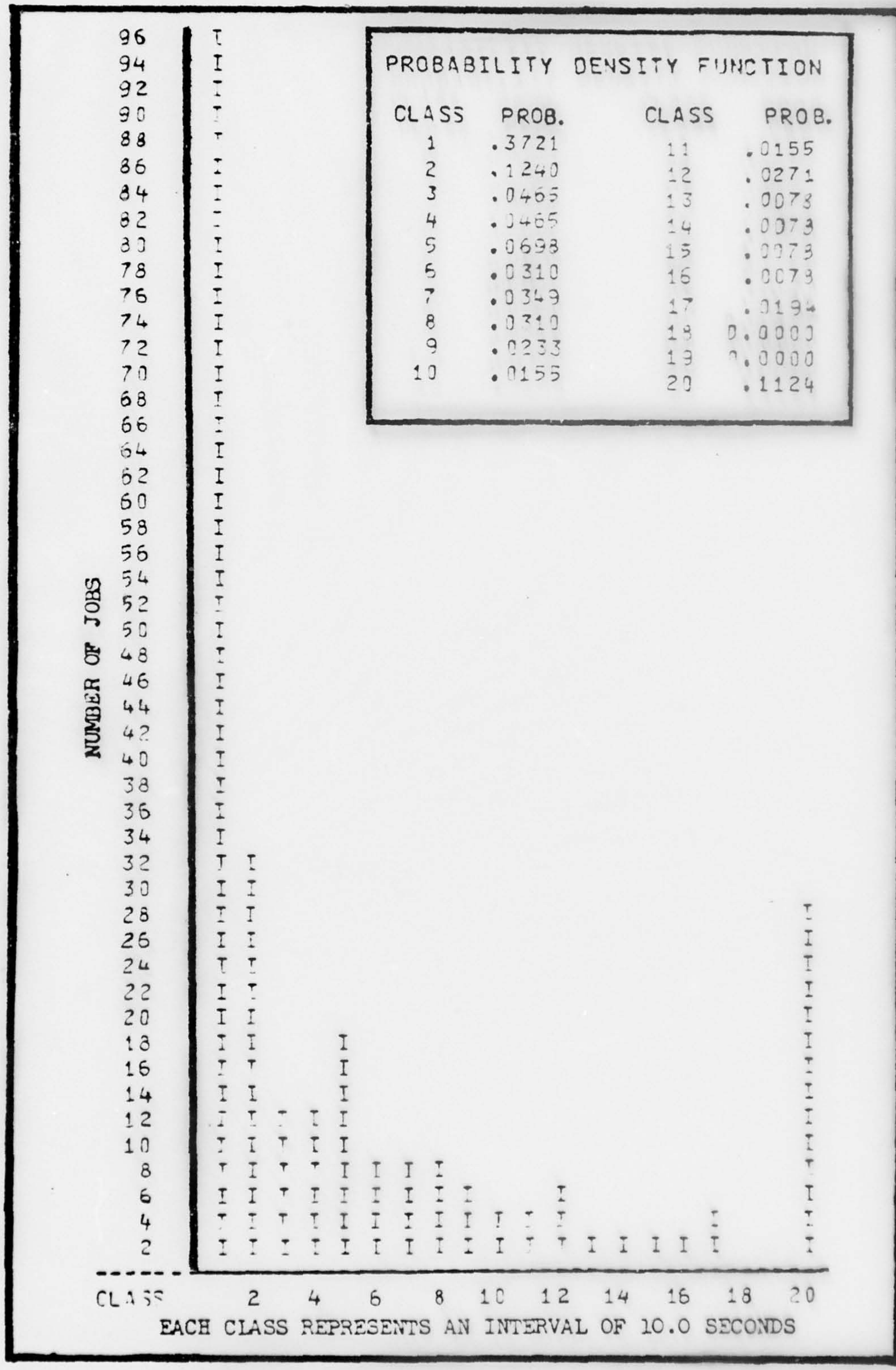


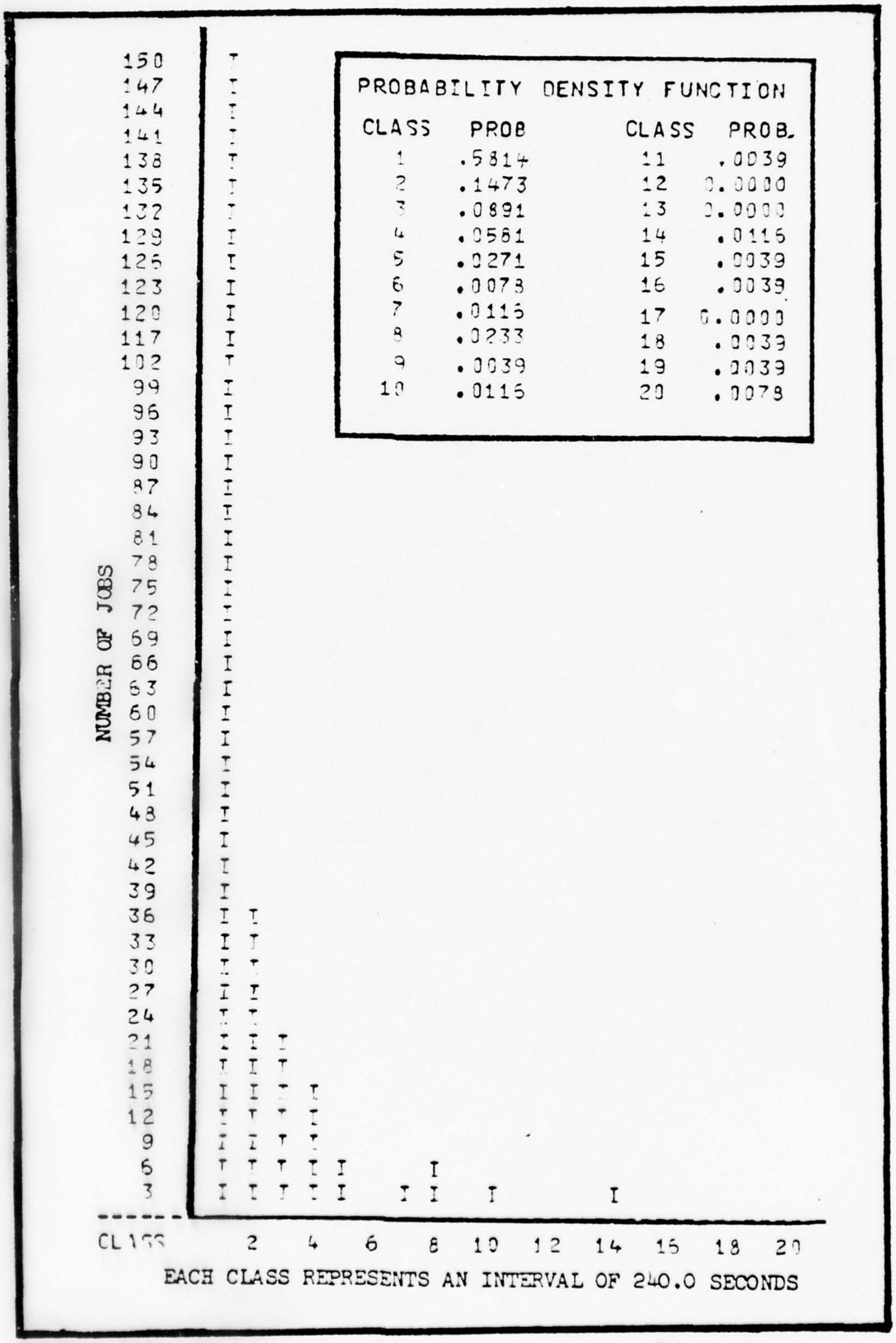
Fig. 33G. Distribution of Control Point Locations for 25 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.3721	11	.0155
2	.1240	12	.0271
3	.0465	13	.0078
4	.0465	14	.0078
5	.0698	15	.0078
6	.0310	16	.0078
7	.0349	17	.0194
8	.0310	18	0.0000
9	.0233	19	0.0000
10	.0155	20	.1124

Fig. 33H. Distribution of Interarrival time for 25 October



CLASS	PROB	CLASS	PROB.
1	.5814	11	.0039
2	.1473	12	0.0000
3	.0891	13	0.0000
4	.0581	14	.0116
5	.0271	15	.0039
6	.0078	16	.0039
7	.0116	17	0.0000
8	.0233	18	.0039
9	.0039	19	.0039
10	.0116	20	.0078

Fig. 33I. Distribution of Input Queue time for 25 October

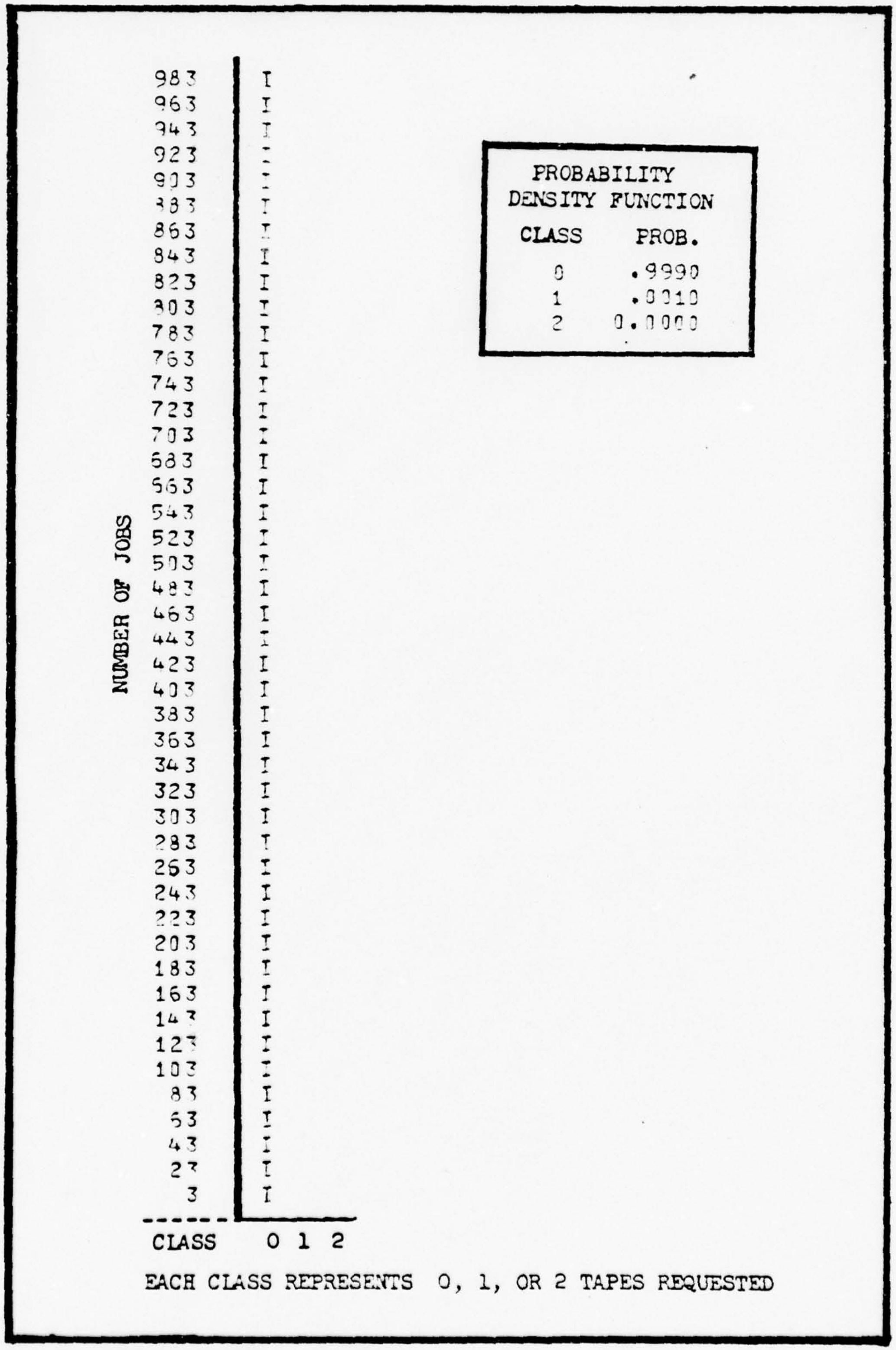


Fig. 33J. Distribution of Tapes Requested for 25 October

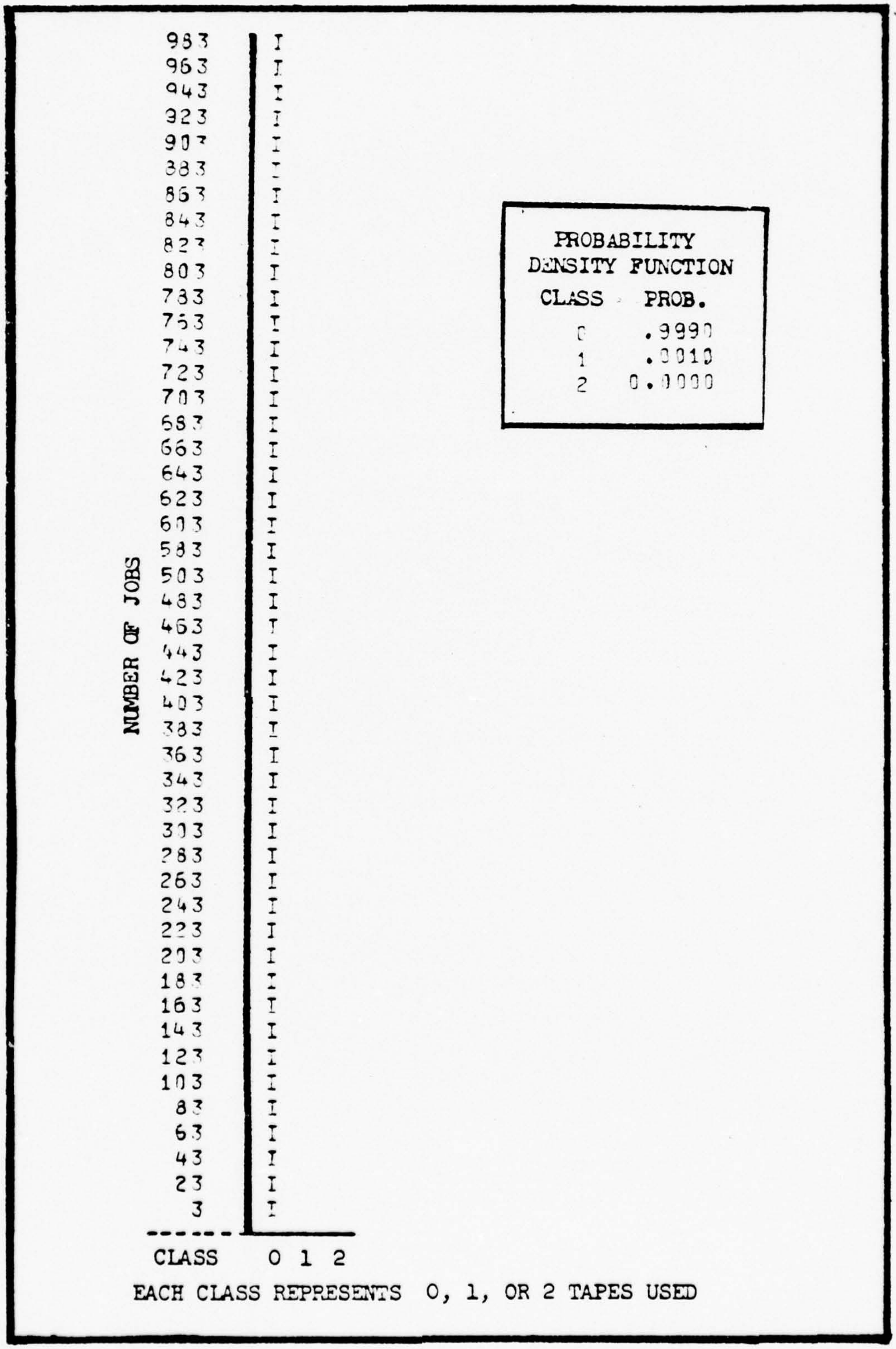


Fig. 33K. Distribution of Tapes Used for 25 October

TABLE XVIII

Statistical Summary for Workload Parameters,

25 October

Variable CPTIME CPU time in seeonds

Mean	24.281	Std Err	1.937	Std Dev	61.423
Variance	3772.805	Kurtosis	132.653	Skewness	9.012
Minimum	.001	Maximum	1130.793	Sum	24426.434
C.V. Pct	252.971	.95 C.I.	20.481	to	28.081

Variable PPTIME Peripheral Processor time in seconds

Mean	78.360	Std Err	42.325	Std Dev	1342.455
Variance	1802186.533	Kurtosis	977.560	Skewness	31.140
Minimum	.481	Maximum	42364.018	Sum	78830.468
C.V. Pct	1713.183	.95 C.I.	-4.696	to	161.417

Variable TIMEIO Input-Output time in seconds

Mean	13.953	Std Err	1.549	Std Dev	49.131
Variance	2413.881	Kurtosis	181.646	Skewness	11.251
Minimum	.074	Maximum	1001.031	Sum	14036.749
C.V. Pct	352.119	.95 C.I.	10.913	to	16.993

Variable TOTCOST Total Cost in CRUs

Mean	24.085	Std Err	1.959	Std Dev	58.926
Variance	3472.300	Kurtosis	65.333	Skewness	6.695
Minimum	.040	Maximum	872.141	Sum	24229.503
C.V. Pct	244.659	.95 C.I.	20.439	to	27.731

TABLE XVIII (continued)

Variable KWS Memory in kilo-word seconds

Mean	648.558	Std Err	63.392	Std Dev	2010.625
Variance	4042614.325	Kurtosis	144.187	Skewness	9.753
Minimum	0.000	Maximum	38459.912	Sum	652449.437
C.V. Pct	310.015	.95 C.I.	524.163	to	772.953

Variable CPCT Control Point Occupancy time in seconds

Mean	1045.156	Std Err	67.905	Std Dev	2153.766
Variance	4638709.153	Kurtosis	36.240	Skewness	4.931
Minimum	1.000	Maximum	25678.000	Sum	1051427.000
C.V. Pct	206.071	.95 C.I.	911.905	to	1178.407

Variable CMLOC Central Memory Locations in Use

Mean	191314.195	Std Err	2922.756	Std Dev	92702.528
Variance	8593E+10	Kurtosis	-.375	Skewness	-1.208
Minimum	512.000	Maximum	261632.000	Sum	.1924E+09
C.V. Pct	48.456	.95 C.I.	185578.791	to	197409.599

Variable CPLOC Control Points in Use

Mean	90.323	Std Err	.571	Std Dev	18.126
Variance	328.559	Kurtosis	10.976	Skewness	-3.320
Minimum	3.000	Maximum	100.000	Sum	90865.000
C.V. Pct	20.068	.95 C.I.	89.202	to	91.445

Variable ROLLOUT Total Time Rolled Out in seconds

Mean	22.127	Std Err	6.715	Std Dev	212.970
Variance	45356.049	Kurtosis	264.394	Skewness	14.786
Minimum	0.000	Maximum	4670.000	Sum	22260.000
C.V. Pct	962.477	.95 C.I.	8.951	to	35.303

TABLE XVIII (continued)

Variable IATIME Inter-arrival time in seconds

Mean	40.063	Std Err	5.439	Std Dev	131.777
Variance	17365.192	Kurtosis	34.576	Skewness	5.512
Minimum	0.000	Maximum	1242.000	Sum	23517.000
C.V. Pct	328.924	.95 C.I.	29.381	to	50.745

Variable INQTIME Time in Input Queue in seconds

Mean	129.744	Std Err	16.680	Std Dev	513.840
Variance	264031.256	Kurtosis	77.905	Skewness	7.645
Minimum	0.000	Maximum	7770.000	Sum	123127.000
C.V. Pct	396.041	.95 C.I.	97.000	to	162.478

Variable TAPEREQ Number of Tapes Requested

Mean	.052	Std Err	.007	Std Dev	.226
Variance	.051	Kurtosis	17.958	Skewness	4.306
Minimum	0.000	Maximum	2.000	Sum	52.000
C.V. Pct	437.141	.95 C.I.	.038	to	.066

Variable TAPEUSED Number of Tapes Used

Mean	.052	Std Err	.007	Std Dev	.226
Variance	.051	Kurtosis	17.958	Skewness	4.306
Minimum	0.000	Maximum	2.000	Sum	52.000
C.V. Pct	437.141	.95 C.I.	.038	to	.066

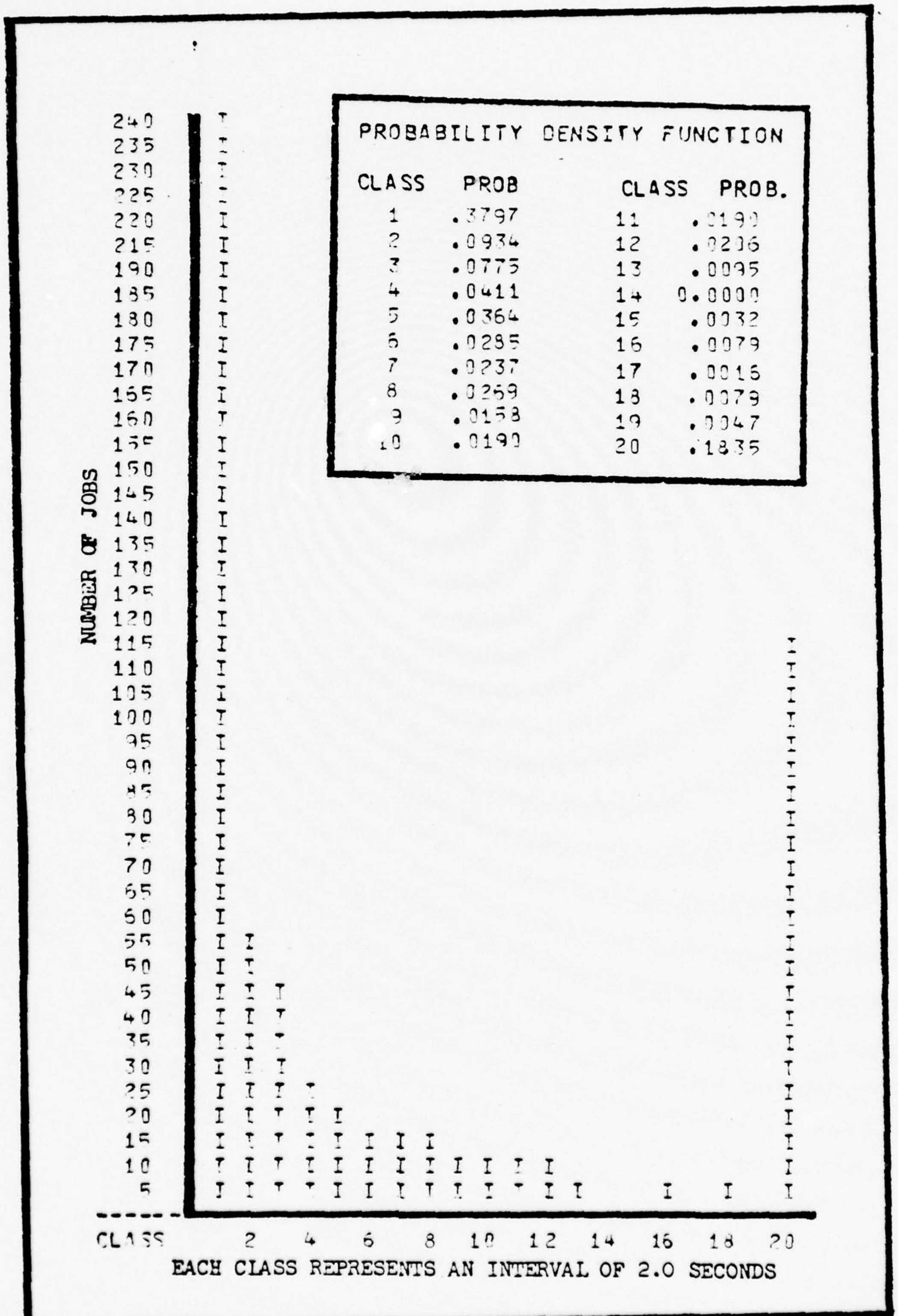
Appendix E

Workload Characterization

30 September

The following variables are characterized with a frequency histogram, probability density function, and/or a statistical summary. All measurements of time are in seconds.

Name	Description
CPTIME	The central processor time needed to process a job.
PPTIME	The peripheral processor time needed to process a job.
TIMEIO	The total I/O time needed to process a job.
TOTCOST	The total job cost in terms of CRUs.
KSW	The memory usage in terms of kilo-word seconds.
CPOT	The total control point occupancy time which is the time a job enters a control point until it leaves a control point for the last time.
CMLOC	The number of central memory locations occupied by all jobs including this job at the time this job left a control point.
CPLOC	The number of control points occupied by all jobs including this job at the time this job left a control point.
ROLLOUT	The total time that a batch job was rolled out for processing magnetic tapes.
IATIME	The total time between this batch job and the last batch job to arrive into the input queue.
INQTIME	The time a batch job spends in the input queue.
TAPEREQ/TAPEUSED	The number of tapes requested/used per job.



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.3797	11	.0190
2	.0934	12	.0206
3	.0775	13	.0095
4	.0411	14	0.0000
5	.0364	15	.0032
6	.0285	16	.0079
7	.0237	17	.0016
8	.0269	18	.0079
9	.0158	19	.0047
10	.0190	20	.1835

Fig. 34A. Distribution of CPU time for 30 September

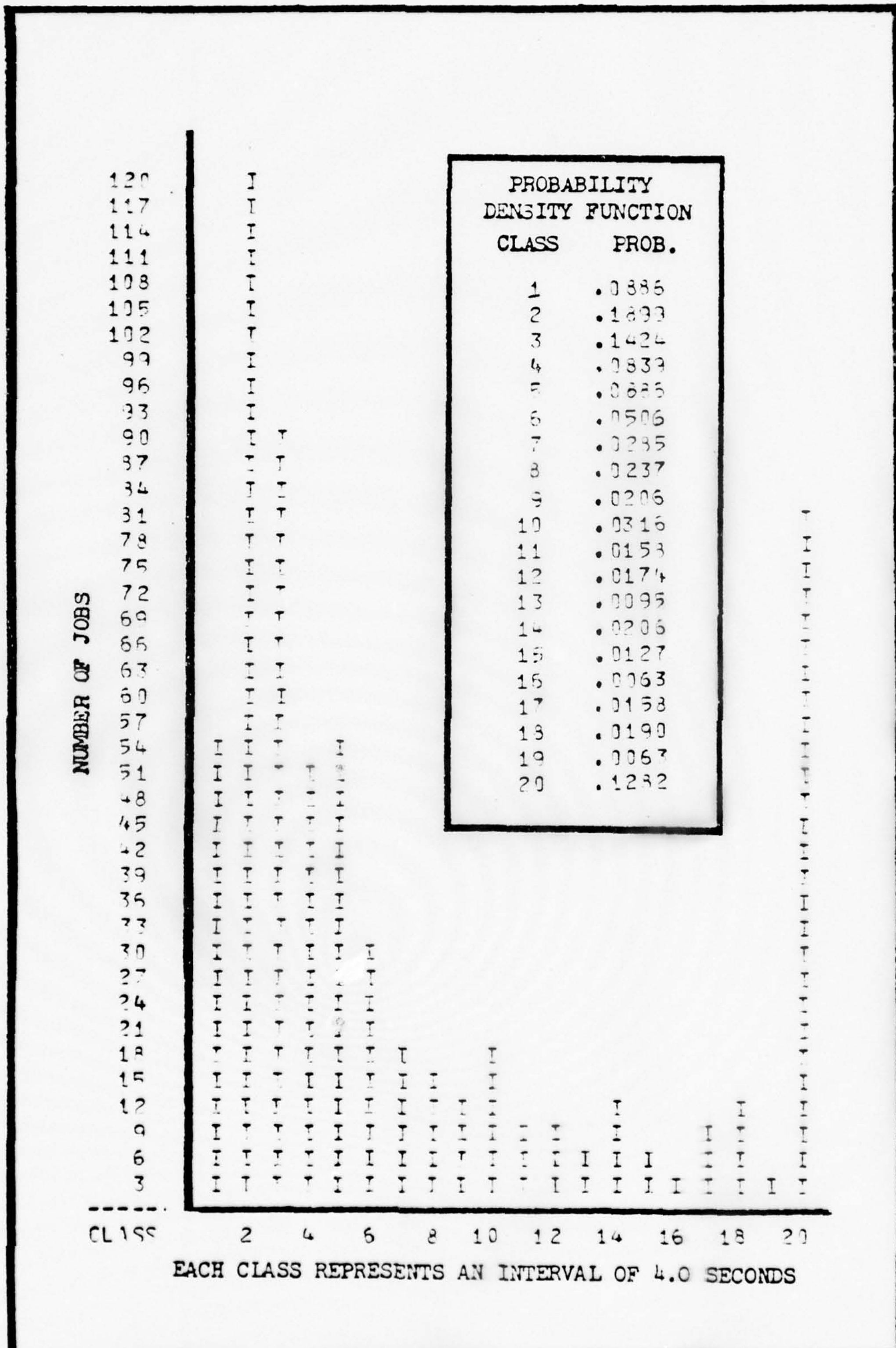
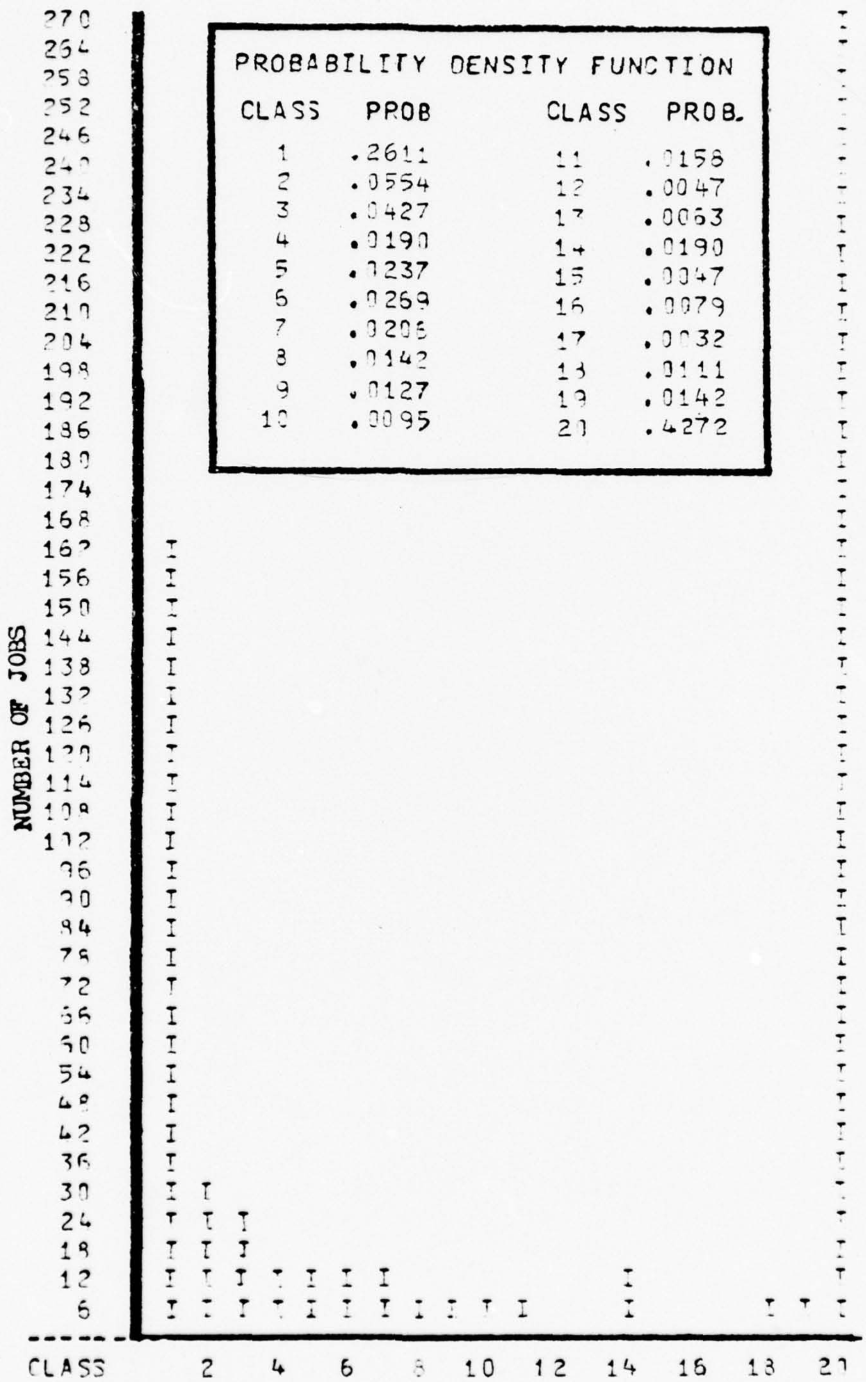


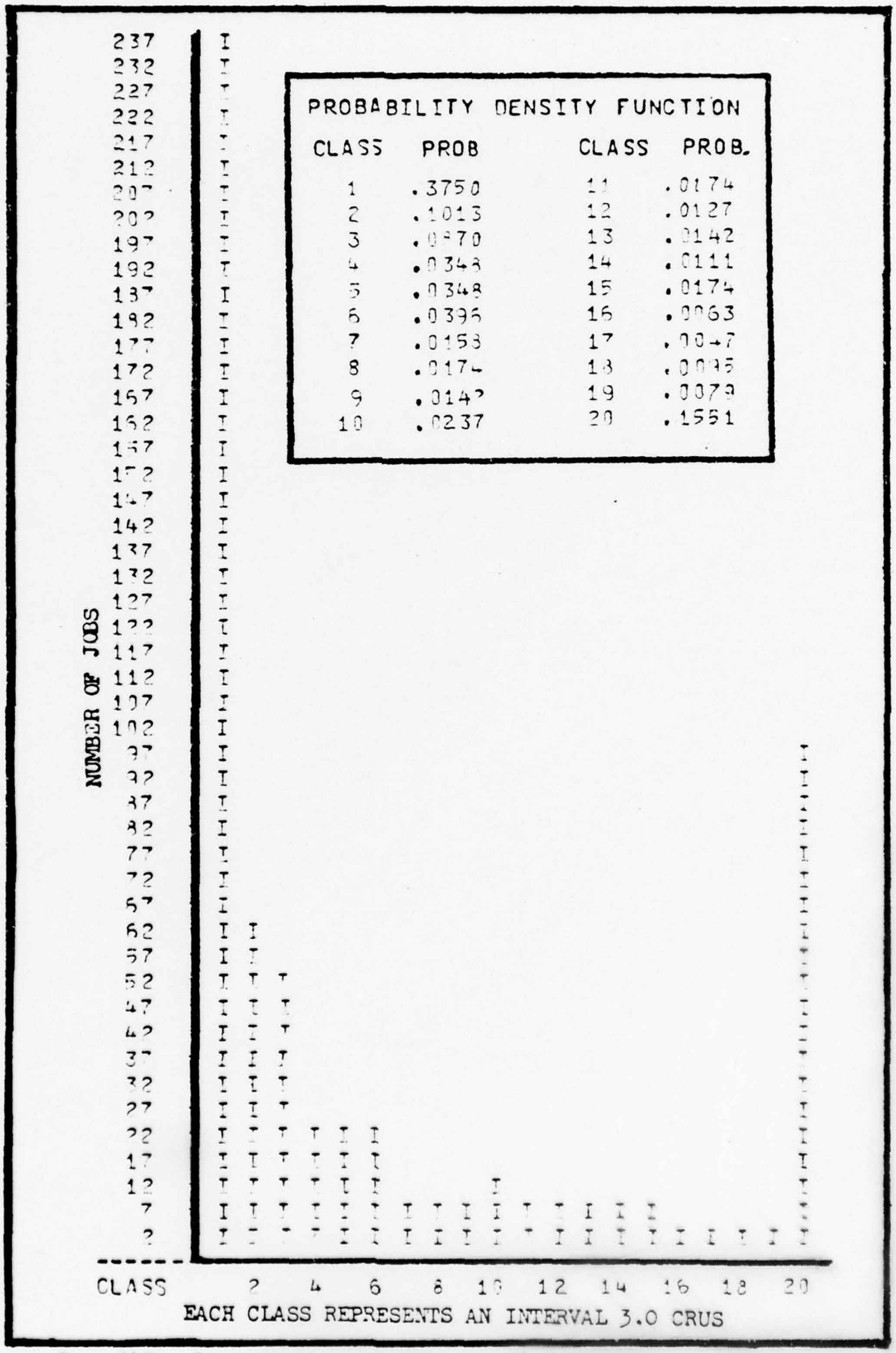
Fig. 34B. Distribution of PPU time for 30 September



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.2611	11	.0158
2	.0554	12	.0047
3	.0427	13	.0063
4	.0190	14	.0190
5	.0237	15	.0047
6	.0269	16	.0079
7	.0206	17	.0032
8	.0142	18	.0111
9	.0127	19	.0142
10	.0095	20	.4272

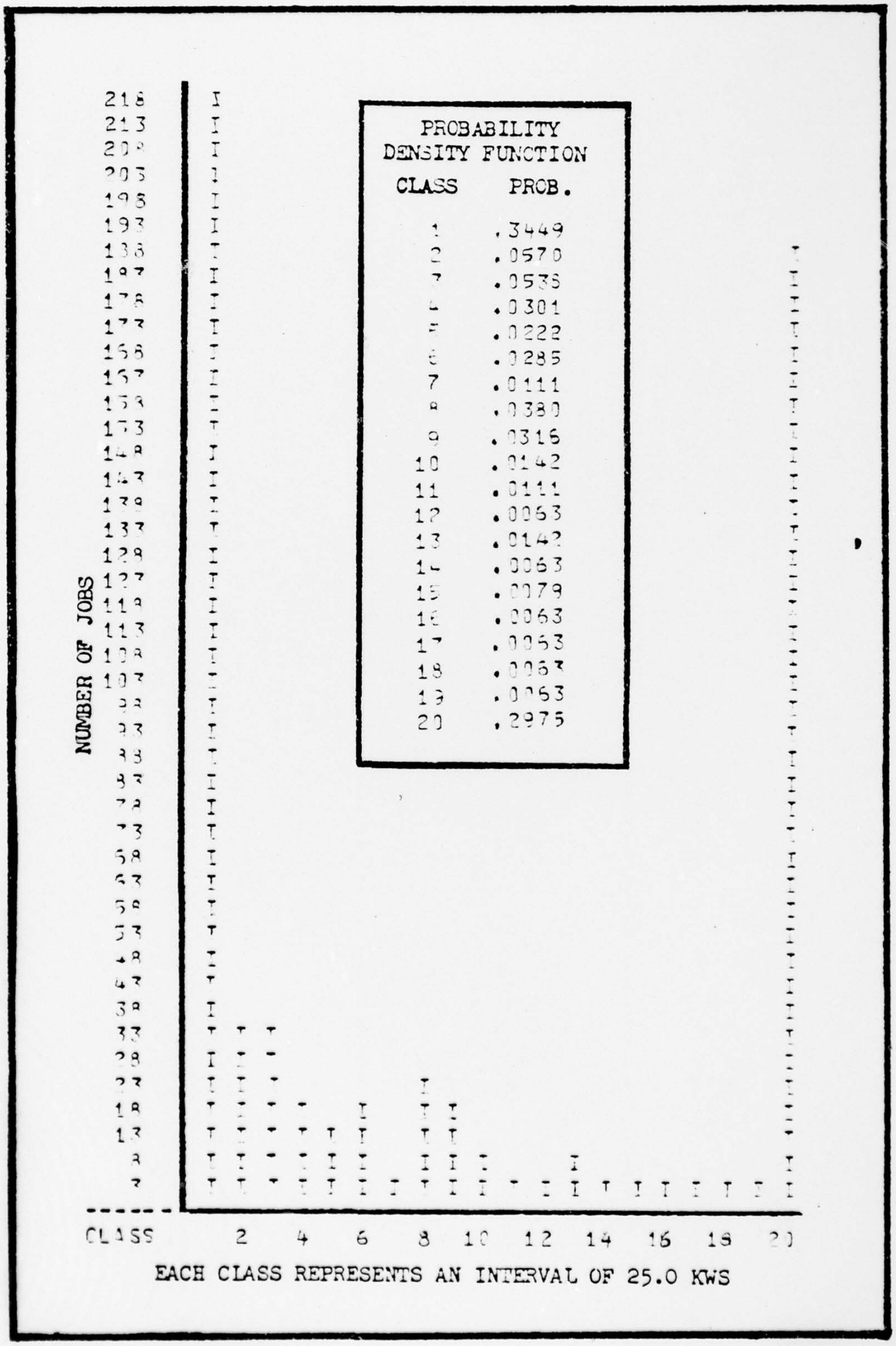
Fig. 34C. Distribution of I/O time for 30 September



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.3750	11	.0174
2	.1013	12	.0127
3	.0870	13	.0142
4	.0348	14	.0111
5	.0348	15	.0174
6	.0396	16	.0063
7	.0158	17	.0047
8	.0174	18	.0095
9	.0142	19	.0079
10	.0237	20	.1551

Fig. 34D. Distribution of CRUs for 30 September



PROBABILITY DENSITY FUNCTION

CLASS	PROB.
1	.3449
2	.0570
3	.0538
4	.0301
5	.0222
6	.0285
7	.0111
8	.0380
9	.0316
10	.0142
11	.0111
12	.0063
13	.0142
14	.0063
15	.0079
16	.0063
17	.0063
18	.0063
19	.0063
20	.2975

Fig. 34E. Distribution of KWS for 30 September

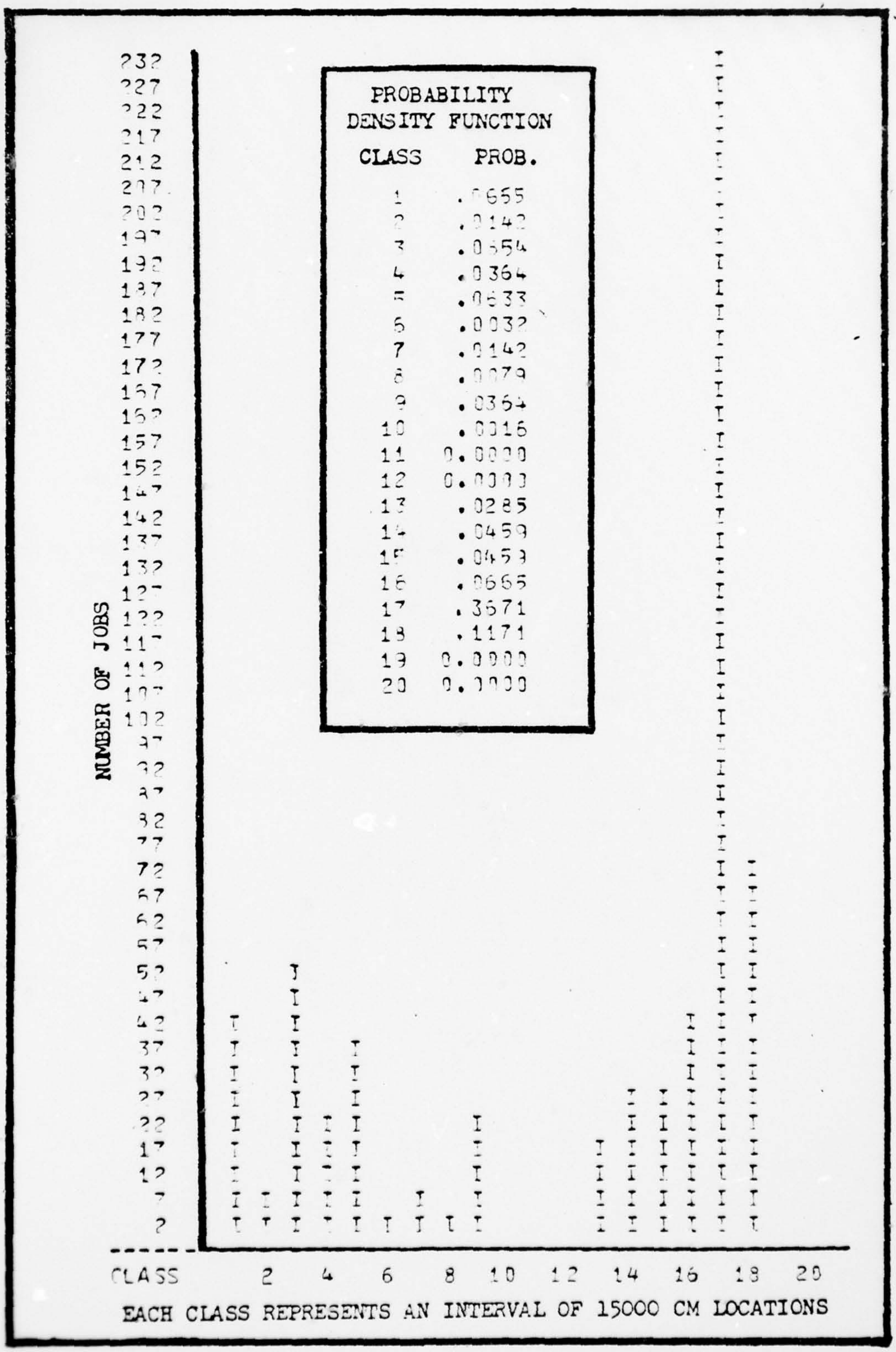


Fig. 34F. Distribution of Central Memory Locations for 30 September

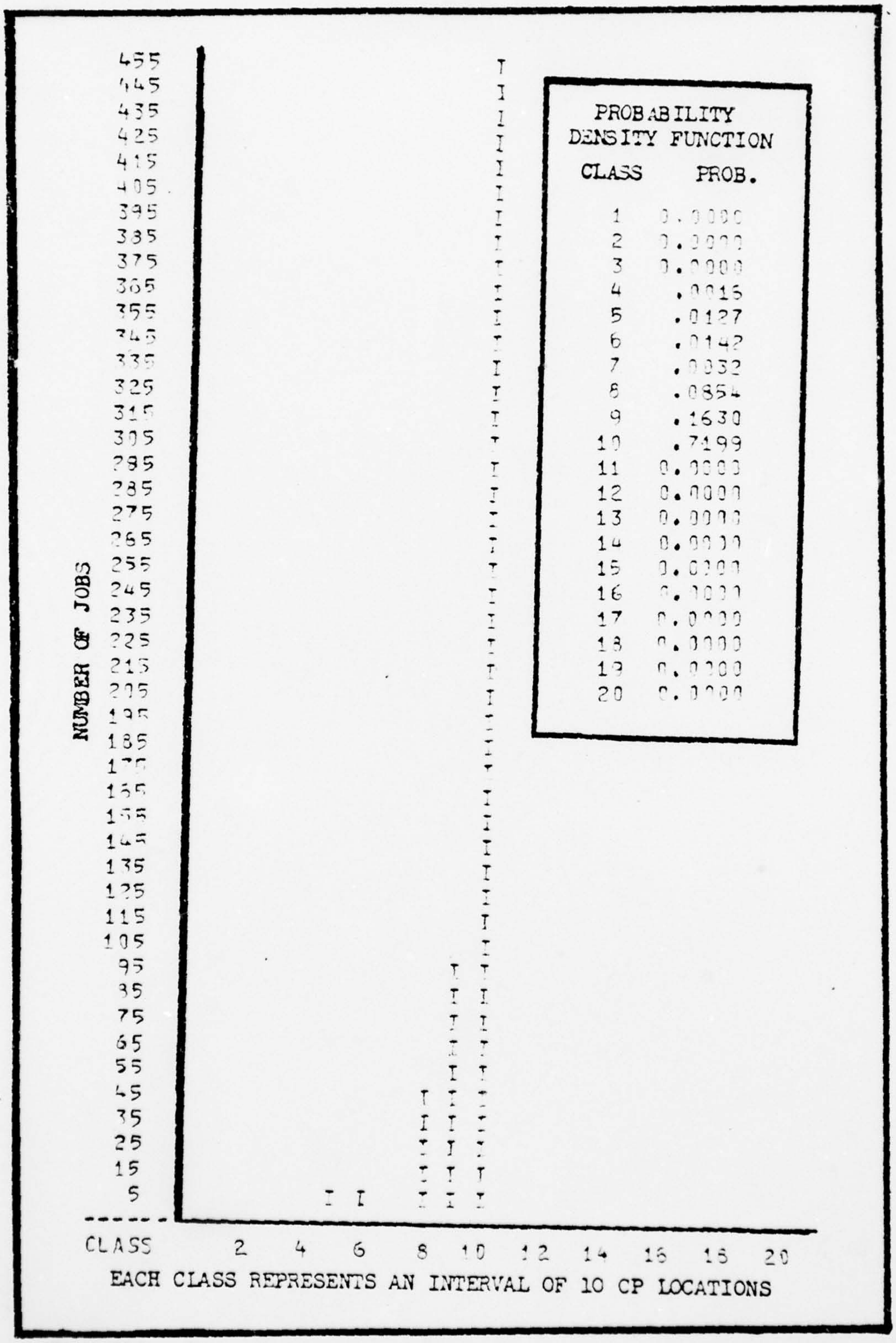
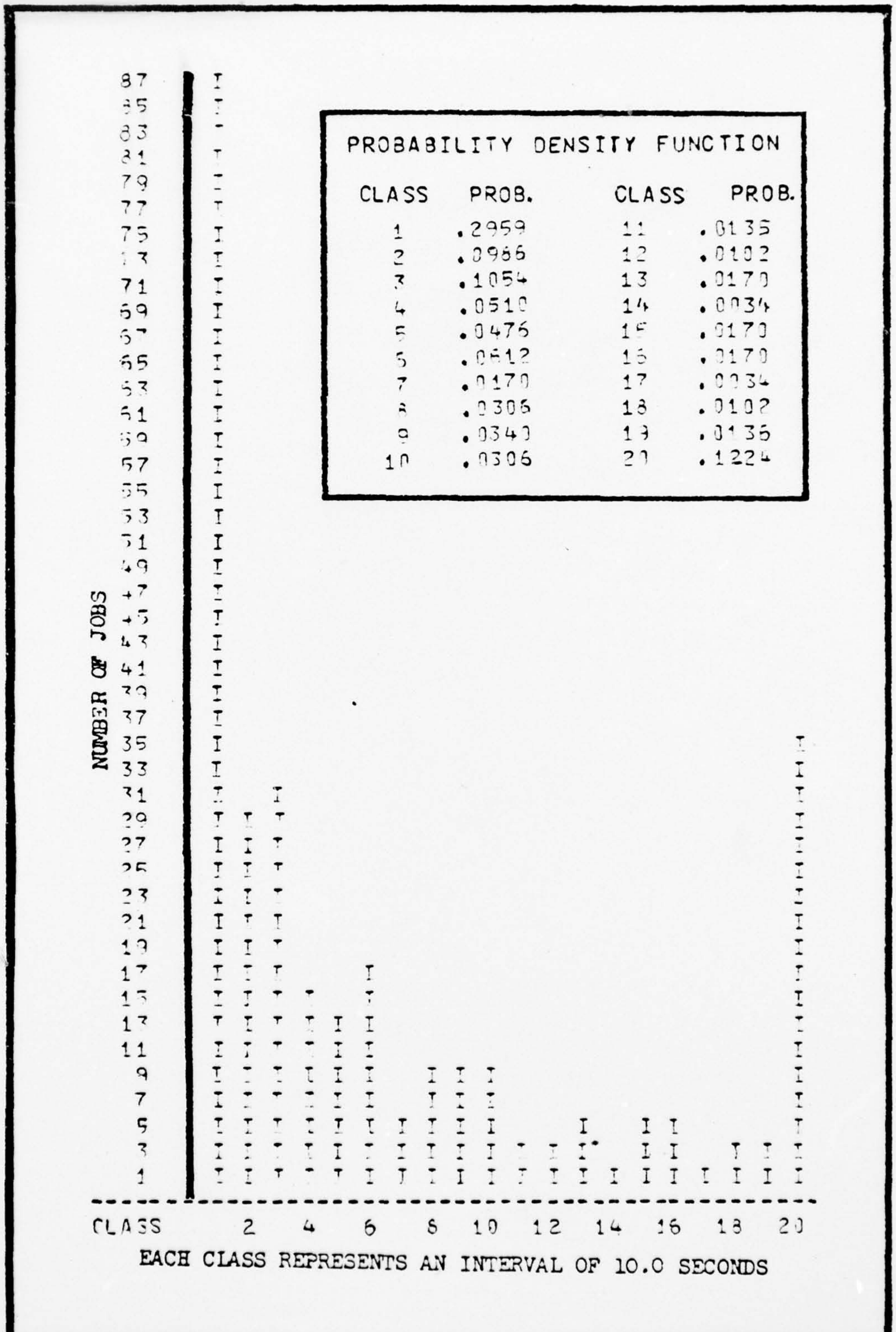


Fig. 34G. Distributions of Control Point Locations for 30 September



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.2959	11	.0136
2	.0986	12	.0102
3	.1054	13	.0170
4	.0510	14	.0034
5	.0476	15	.0170
6	.0612	16	.0170
7	.0170	17	.0034
8	.0306	18	.0102
9	.0340	19	.0136
10	.0306	20	.1224

Fig. 34H. Distribution of Interarrival time for 30 September

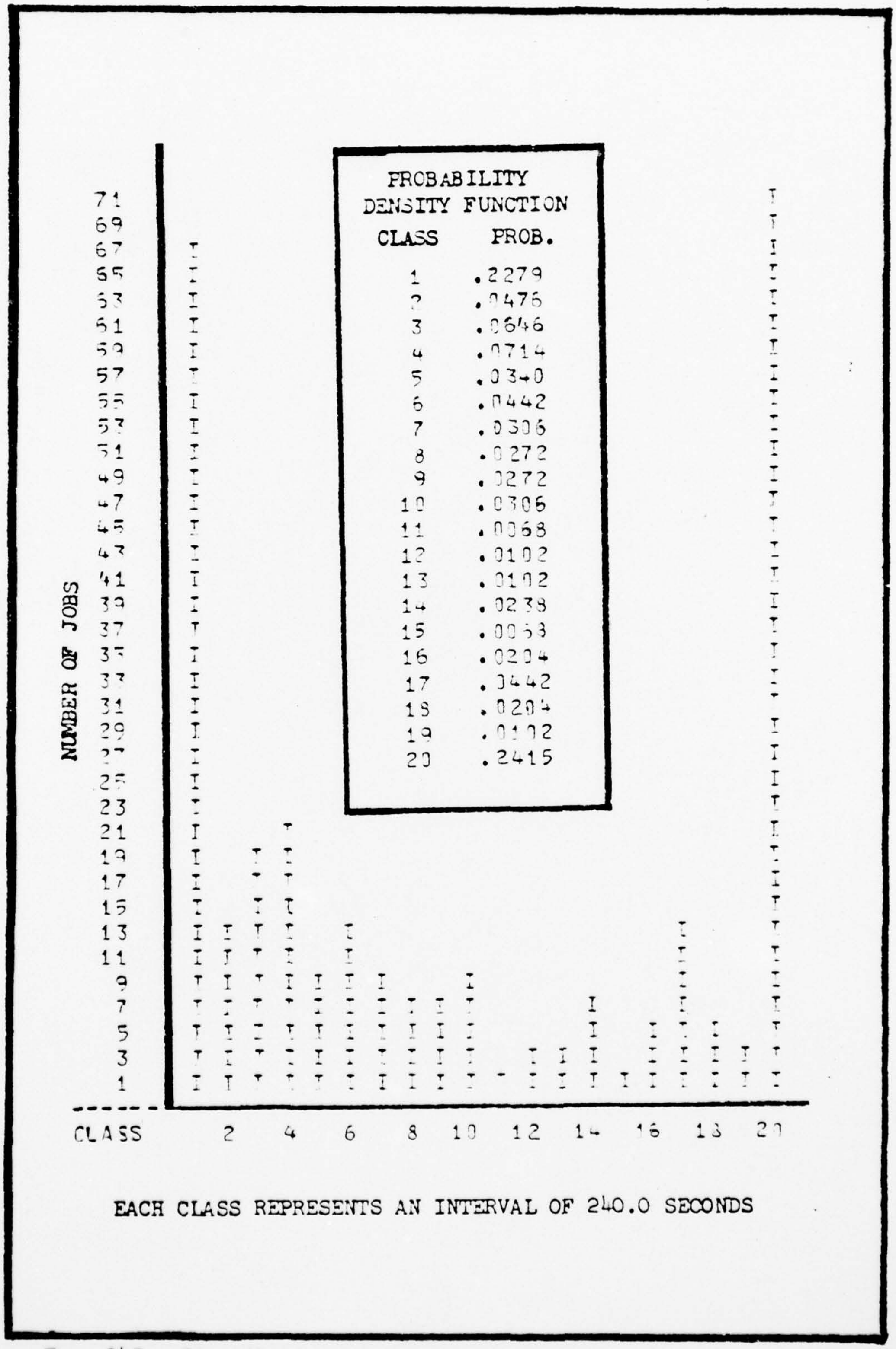


Fig. 34I. Distribution of Input Queue time for 30 September

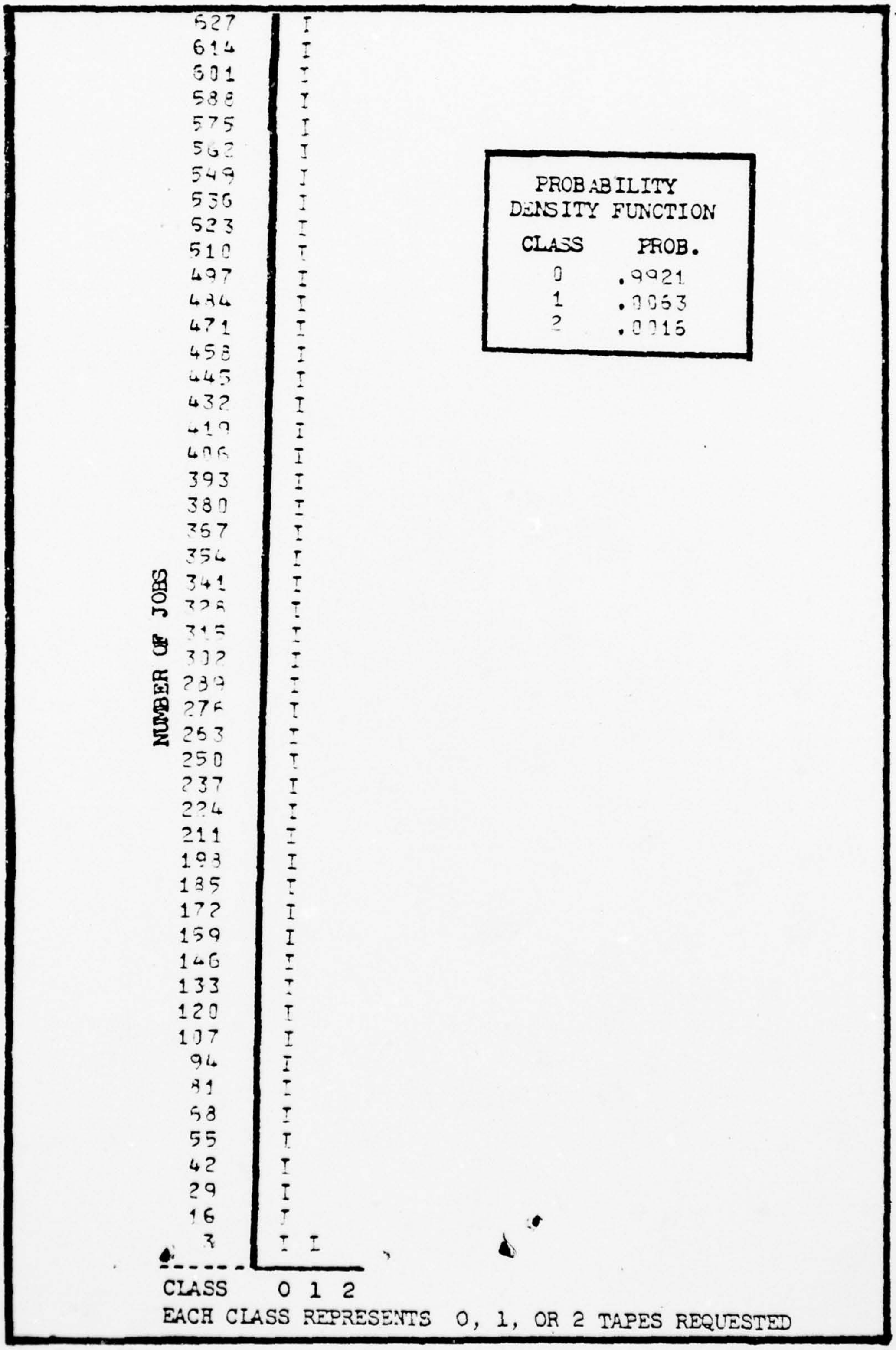


Fig. 34J. Distribution of Tapes Requested for 30 September

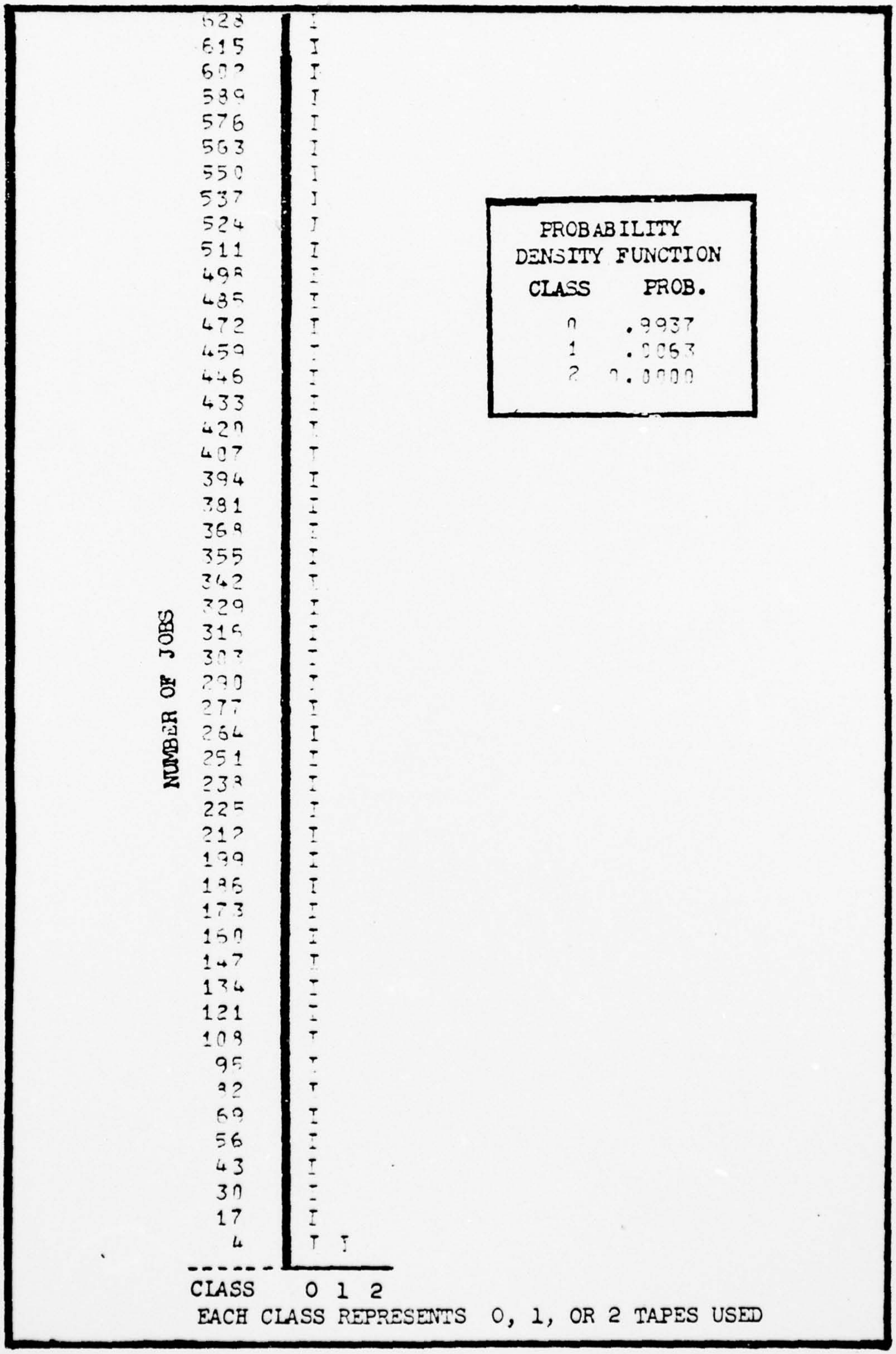


Fig. 34K. Distribution of Tapes Used for 30 September

TABLE XIX

Statistical Summary for Workload Parameters

30 September

Variable CPTIME CPU time in seconds

Mean	29.146	Std Err	2.993	Std Dev	75.647
Variance	5722.529	Kurtosis	87.960	Skewness	7.578
Minimum	.001	Maximum	1148.798	Sum	18623.985
C.V. Pct	259.551	.95 C.I.	23.269	to	35.022

Variable PPTIME Peripheral Processor time in seconds

Mean	83.757	Std Err	26.129	Std Dev	660.507
Variance	436268.919	Kurtosis	318.257	Skewness	16.981
Minimum	.731	Maximum	13814.643	Sum	53520.550
C.V. Pct	788.601	.95 C.I.	32.447	to	135.066

Variable TIMEIO Input-Output time in seconds

Mean	1056.661	Std Err	181.448	Std Dev	4586.719
Variance	.2103E+08	Kurtosis	192.469	Skewness	12.763
Minimum	.223	Maximum	80639.576	Sum	675206.694
C.V. Pct	43.076	.95 C.I.	700.354	to	1412.969

Variable TOTCOST Total Cost in CRUs

Mean	36.515	Std Err	5.461	Std Dev	138.036
Variance	19054.035	Kurtosis	280.604	Skewness	14.849
Minimum	.044	Maximum	2859.326	Sum	23333.113
C.V. Pct	358.026	.95 C.I.	25.792	to	47.238

TABLE XIX (continued)

Variable KWS Memory in kilo-word seconds

Mean	1056.699	Std Err	181.463	Std Dev	4587.101
Variance	2104E+08	Kurtosis	192.469	Skewness	12.763
Minimum	.167	Maximum	80646.241	Sum	675230.971
C.V. Pct	434.097	.95 C.I.	700.363	to	1413.036

Variable CPCT Control Point Occupancy time in seconds

Mean	1103.853	Std Err	78.746	Std Dev	1990.586
Variance	3962434.217	Kurtosis	13.954	Skewness	3.356
Minimum	1.000	Maximum	15111.000	Sum	705362.000
C.V. Pct	180.331	.95 C.I.	949.219	to	1258.486

Variable CMLOC Central Memory Locations in Use

Mean	180052.532	Std Err	3606.704	Std Dev	91171.877
Vari	8312E+10	Kurtosis	-1.059	Skewness	-.822
Minimum	512.000	Maximum	261632.000	Sum	.1150E+09
C.V. Pct	50.636	.95 C.I.	172970.067	to	187134.977

Variable CPLOC Control Points in Use

Mean	91.825	Std Err	.422	Std Dev	10.679
Variance	115.048	Kurtosis	5.401	Skewness	-2.114
Minimum	40.000	Maximum	100.000	Sum	58676.000
C.V.	11.630	.95 C.I.	90.995	to	92.654

Variable ROLLOUT Total Time Rolled Out in seconds

Mean	42.147	Std Err	12.004	Std Dev	303.440
Variance	92075.941	Kurtosis	149.443	Skewness	11.383
Minimum	0.000	Maximum	4952.000	Sum	26932.000
C.V. Pct	719.955	.95 C.I.	18.575	to	65.719

TABLE XIX (continued)

Variable IATIME   Inter-arrival time in seconds

Mean	101.364	Std Err	17.755	Std Dev	311.599
Variance	97094.043	Kurtosis	107.143	Skewness	9.758
Minimum	0.000	Maximum	3785.000	Sum	31220.000
C.V. Pct	707.407	.95 C.I.	66.427	to	136.301

Variable INQTIME   Time in Input Queue in seconds

Mean	1533.987	Std Err	122.126	Std Dev	2981.482
Variance	8889235.859	Kurtosis	5.235	Skewness	2.358
Minimum	0.000	Maximum	17387.000	Sum	914[illegible][illegible]6.000
C.V. Pct	194.362	.95 C.I.	1294.136	to	[illegible]773.838

Variable TAPEREQ   Number of Tapes Requested

Mean	.097	Std Err	.013	Std Dev	.331
Variance	.110	Kurtosis	17.540	Skewness	3.845
Minimum	0.000	Maximum	3.000	Sum	62.000
C.V. Pct	341.350	.95 C.I.	.071	to	.123

Variable TAPEUSED   Number of Tapes Used

Mean	.094	Std Err	.012	Std Dev	.313
Variance	.098	Kurtosis	11.441	Skewness	3.382
Minimum	0.000	Maximum	2.000	Sum	60.000
C.V. Pct	332.977	.95 C.I.	.070	to	.118

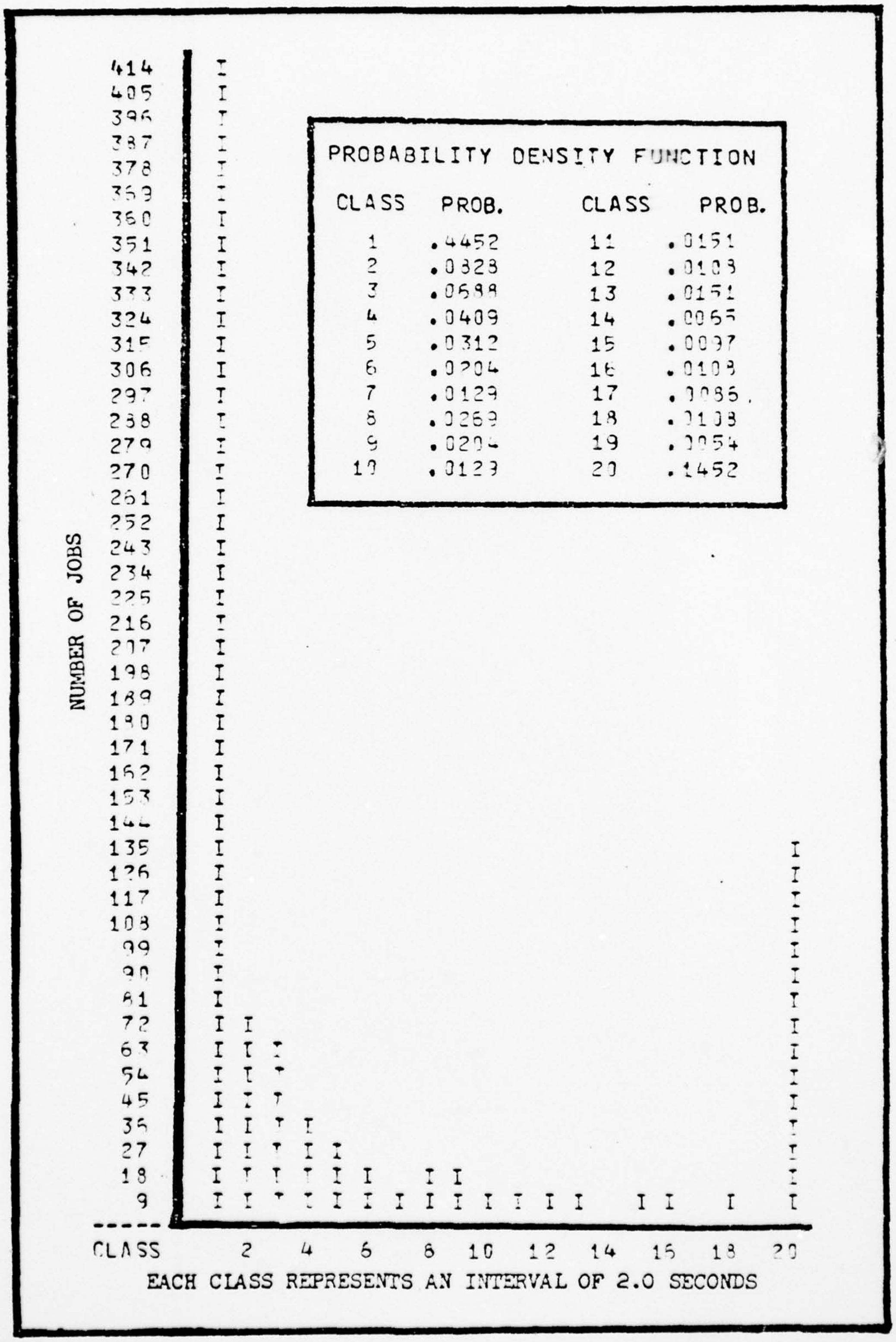
Appendix F

Workload Characterization

14 October

The following variables are characterized with a frequency histogram, probability density function, and/or a statistical summary. All measurements of time are in seconds.

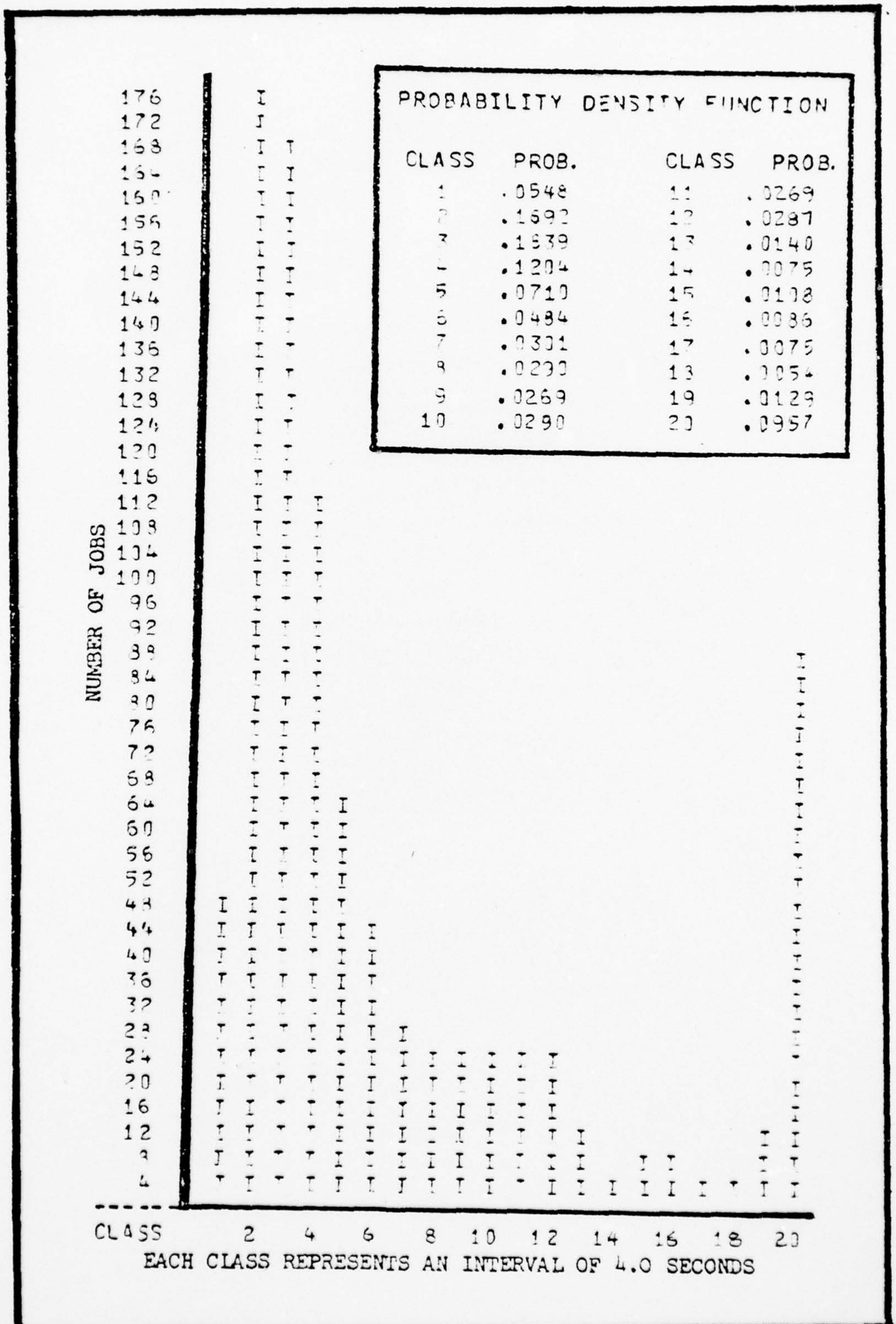
Name	Description
CPTIME	The central processor time needed to process a job.
PPTIME	The peripheral processor time needed to process a job.
TIMEIO	The total I/O time needed to process a job.
TOTCOST	The total job cost in terms of CRUs.
KSW	The memory usage in terms of kilo-word seconds.
CPCT	The total control point occupancy time which is the time a job enters a control point until it leaves a control point for the last time.
CMLOC	The number of central memory locations occupied by all jobs including this job at the time this job left a control point.
CPLOC	The number of control points occupied by all jobs including this job at the time this job left a control point.
ROLLOUT	The total time that a batch job was rolled out for processing magnetic tapes.
IATIME	The total time between this batch job and the last batch job to arrive into the input queue.
INQTIME	The time a batch job spends in the input queue.
TAPEREQ/TAPEUSED	The number of tapes requested/used per job.



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.4452	11	.0151
2	.0828	12	.0108
3	.0688	13	.0151
4	.0409	14	.0065
5	.0312	15	.0097
6	.0204	16	.0108
7	.0129	17	.0086
8	.0269	18	.0108
9	.0204	19	.0054
10	.0129	20	.1452

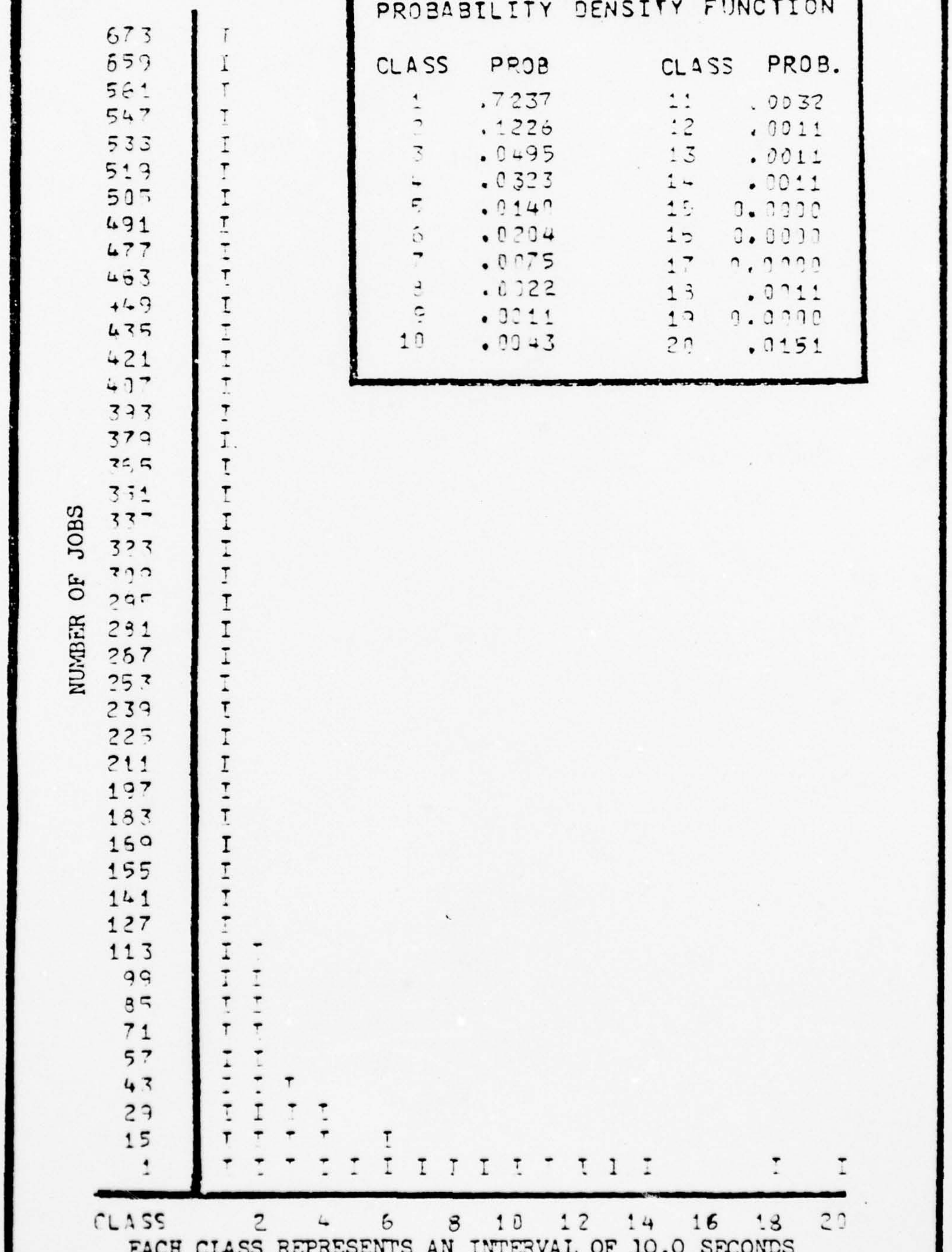
Fig. 35A. Distribution of CPU time for 14 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.0548	11	.0269
2	.1692	12	.0287
3	.1639	13	.0140
4	.1204	14	.0075
5	.0710	15	.0108
6	.0484	16	.0086
7	.0301	17	.0075
8	.0290	18	.0054
9	.0269	19	.0129
10	.0290	20	.0957

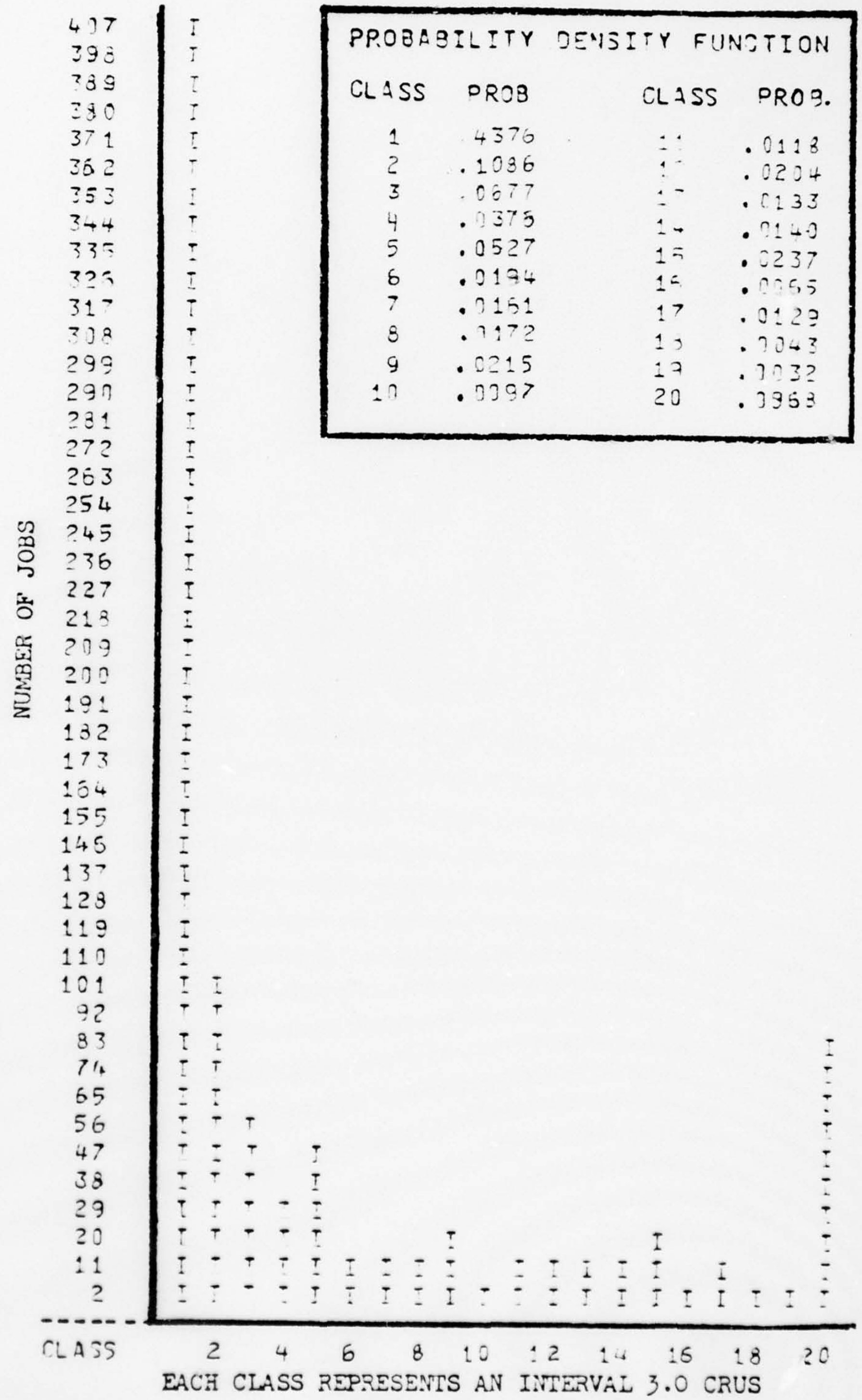
Fig. 35B. Distribution of PPU time for 14 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.7237	11	.0032
2	.1226	12	.0011
3	.0495	13	.0011
4	.0323	14	.0011
5	.0140	15	0.0000
6	.0204	16	0.0000
7	.0075	17	0.0000
8	.0022	18	.0011
9	.0011	19	0.0000
10	.0043	20	.0151

Fig. 35C. Distribution of I/O time for 14 October



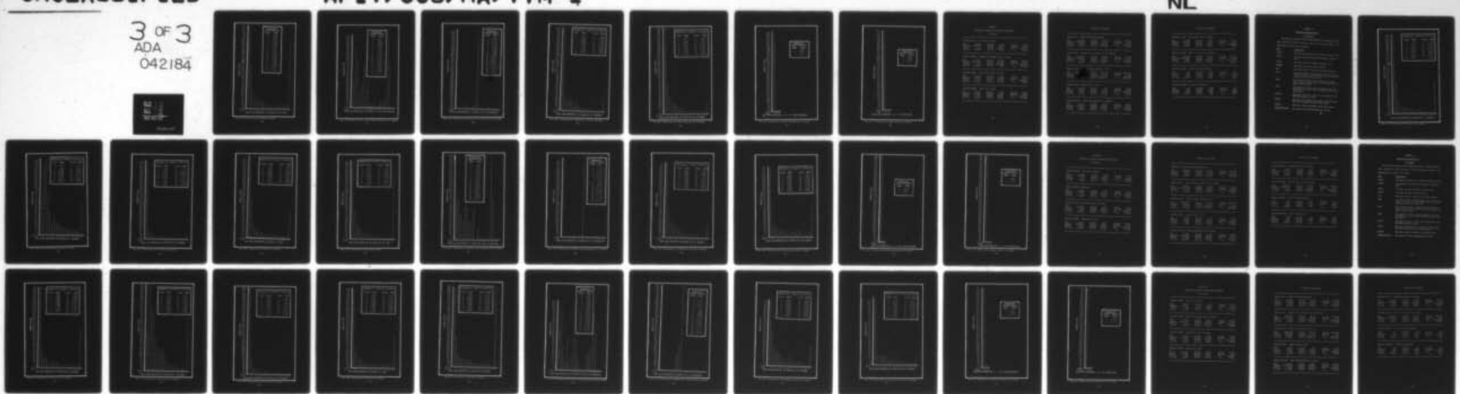
PROBABILITY DENSITY FUNCTION

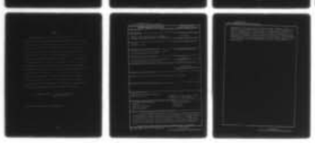
CLASS	PROB	CLASS	PROB.
1	.4376	11	.0118
2	.1086	12	.0204
3	.0677	13	.0133
4	.0375	14	.0140
5	.0527	15	.0237
6	.0194	16	.0065
7	.0161	17	.0129
8	.0172	18	.0043
9	.0215	19	.0032
10	.0097	20	.0968

Fig. 35D. Distribution of CRUs for 14 October

AD-A042 184 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO SCH--ETC F/G 9/2
WORKLOAD CHARACTERIZATION AND MEASUREMENT OF THE CDC CYBER 74 C--ETC(U)
MAR 77 J R BEAR
UNCLASSIFIED AFIT/GCS/MA/77M-1 NL

3 OF 3
ADA
042184





END
DATE
FILMED
8-77

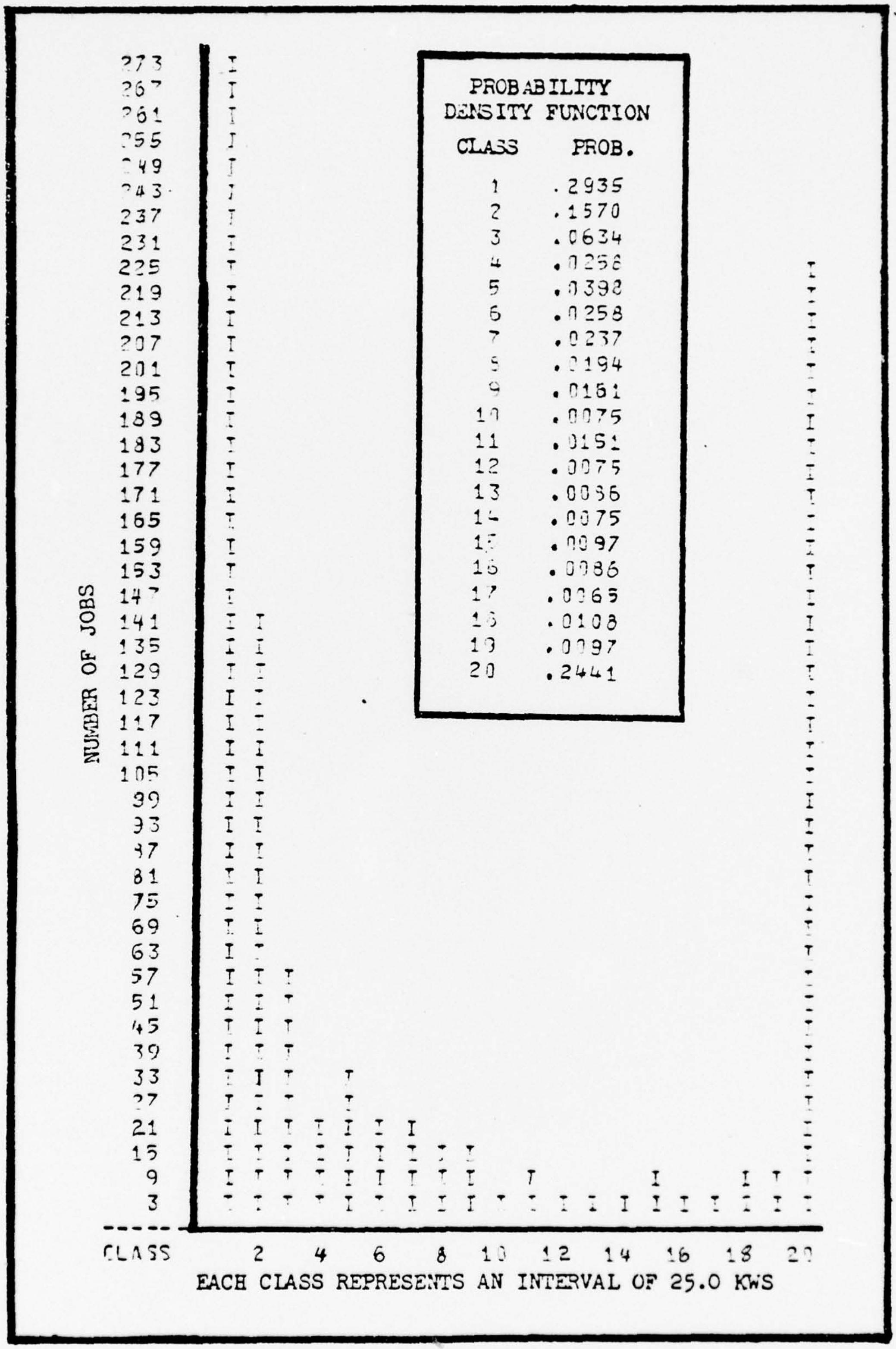


Fig. 35E. Distribution of KWS for 14 October

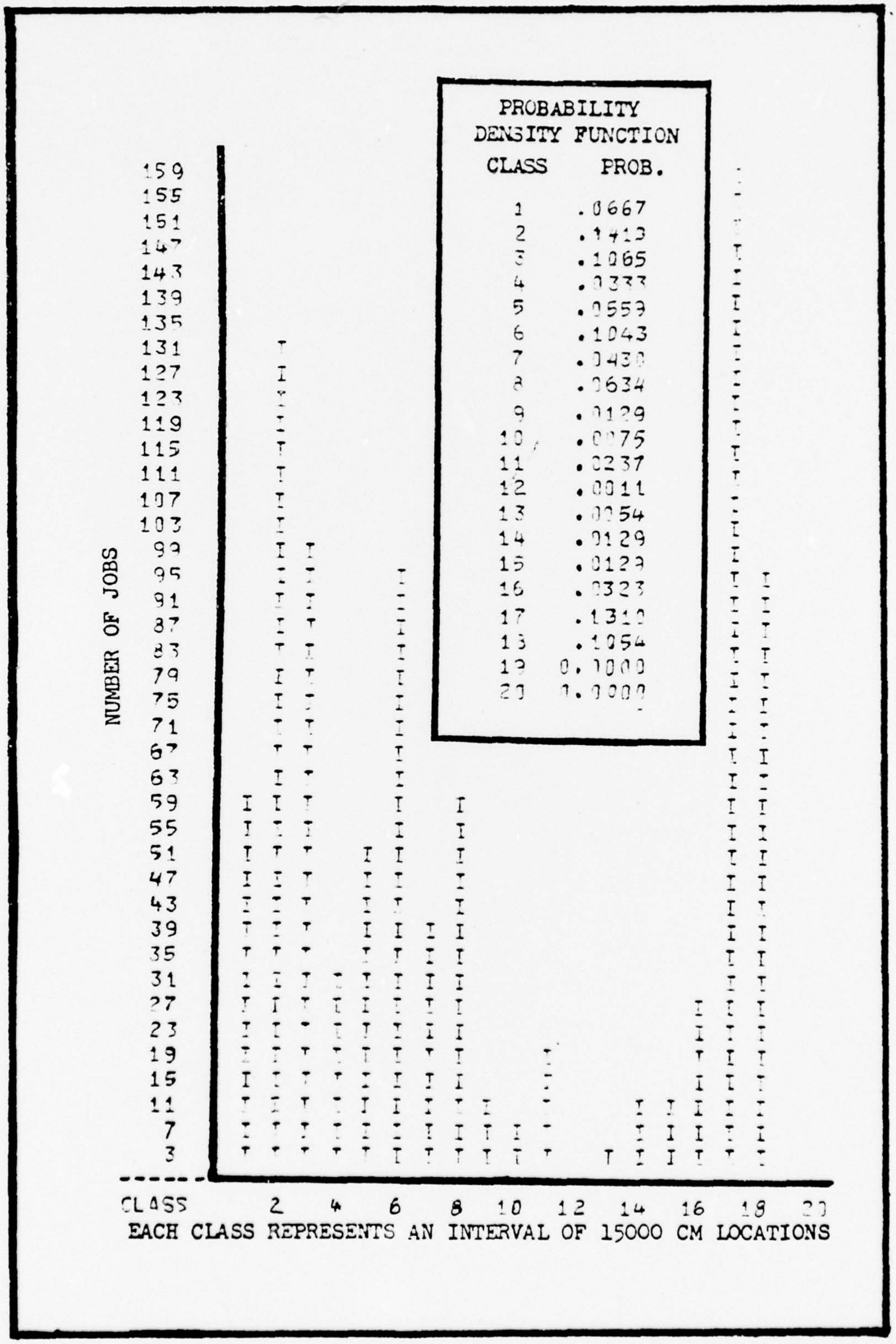
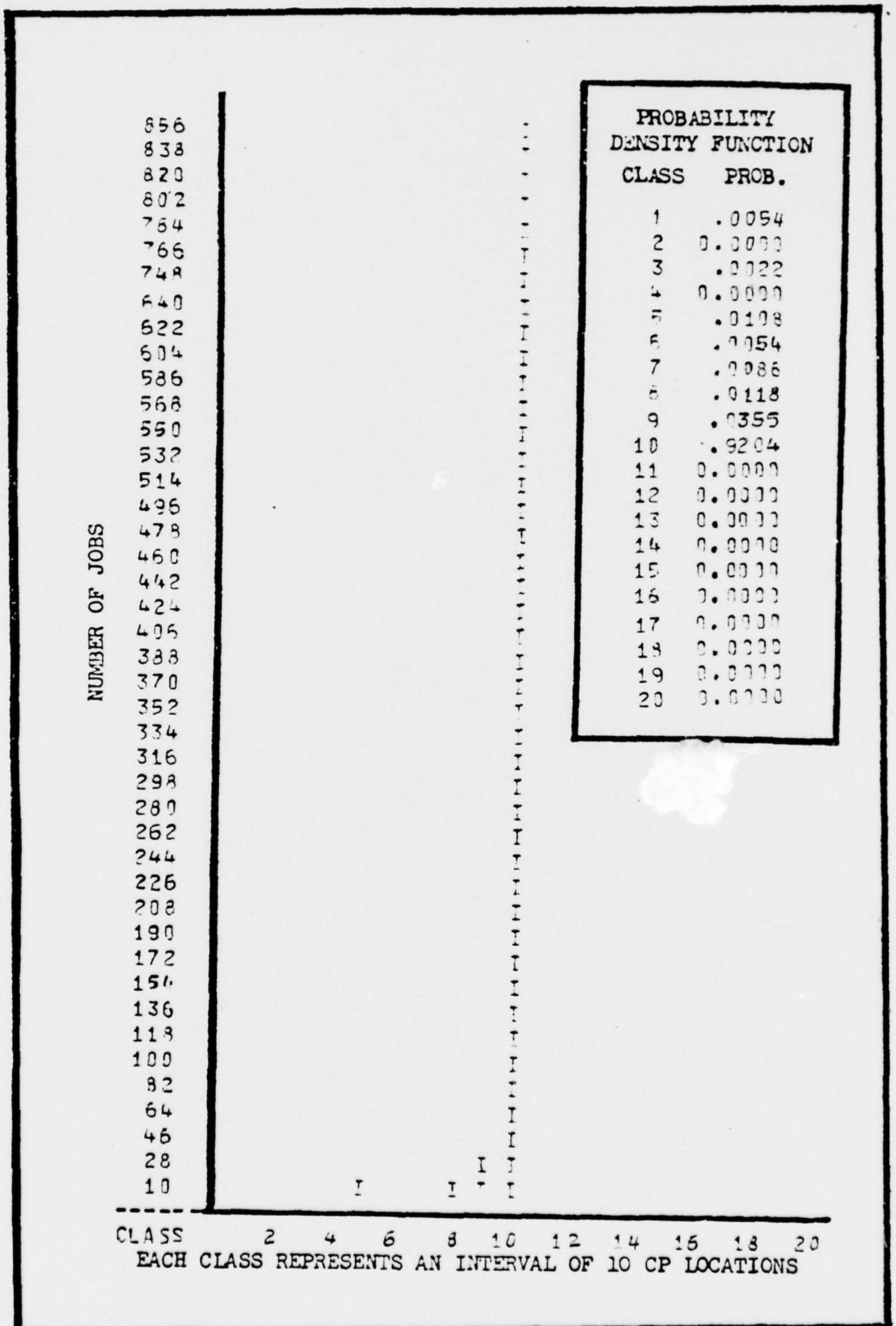


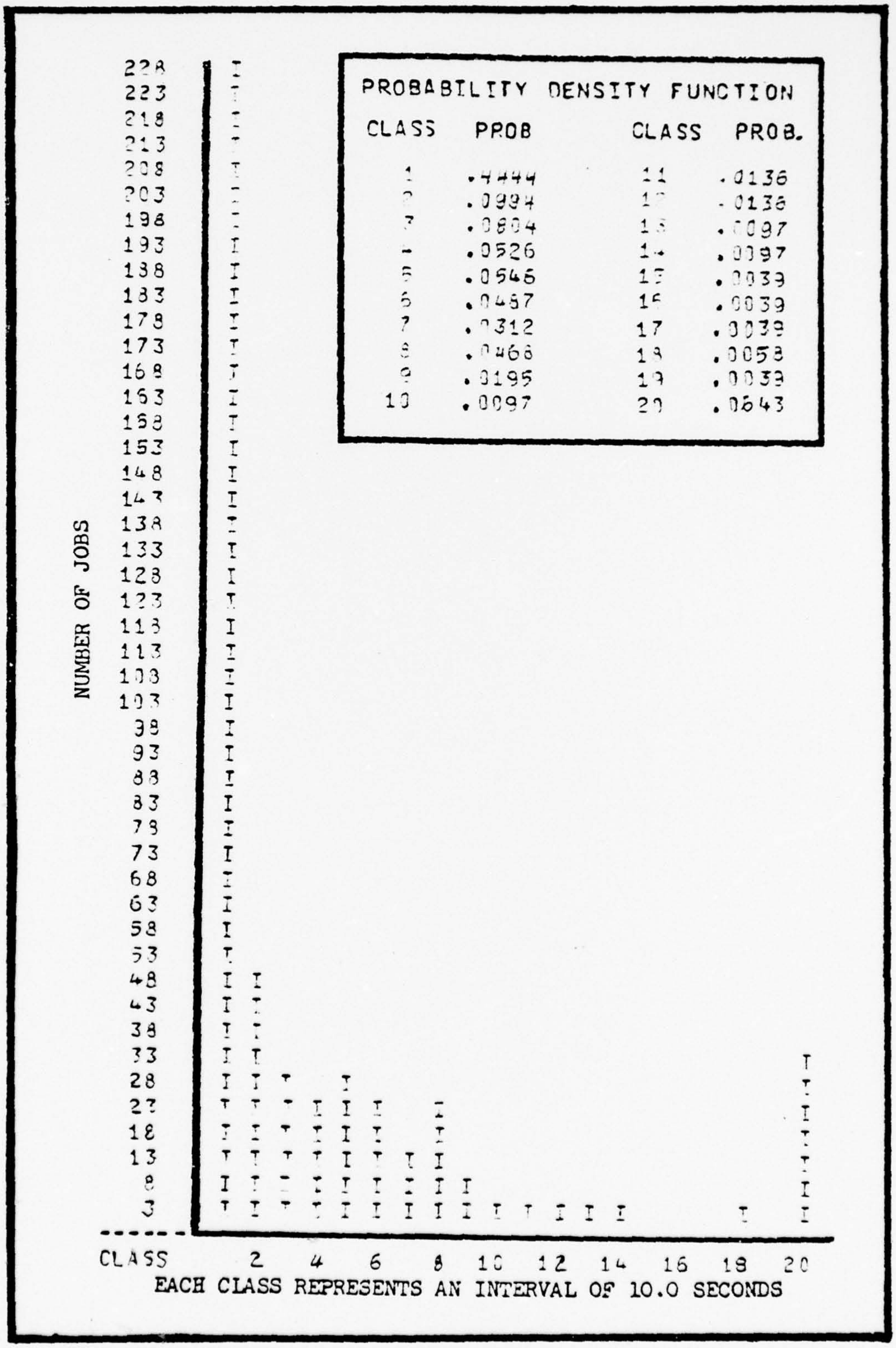
Fig. 35F. Distribution of Central Memory Location for 14 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.
1	.0054
2	0.0000
3	.0022
4	0.0000
5	.0108
6	.0054
7	.0086
8	.0118
9	.0355
10	.9204
11	0.0000
12	0.0000
13	0.0000
14	0.0000
15	0.0000
16	0.0000
17	0.0000
18	0.0000
19	0.0000
20	0.0000

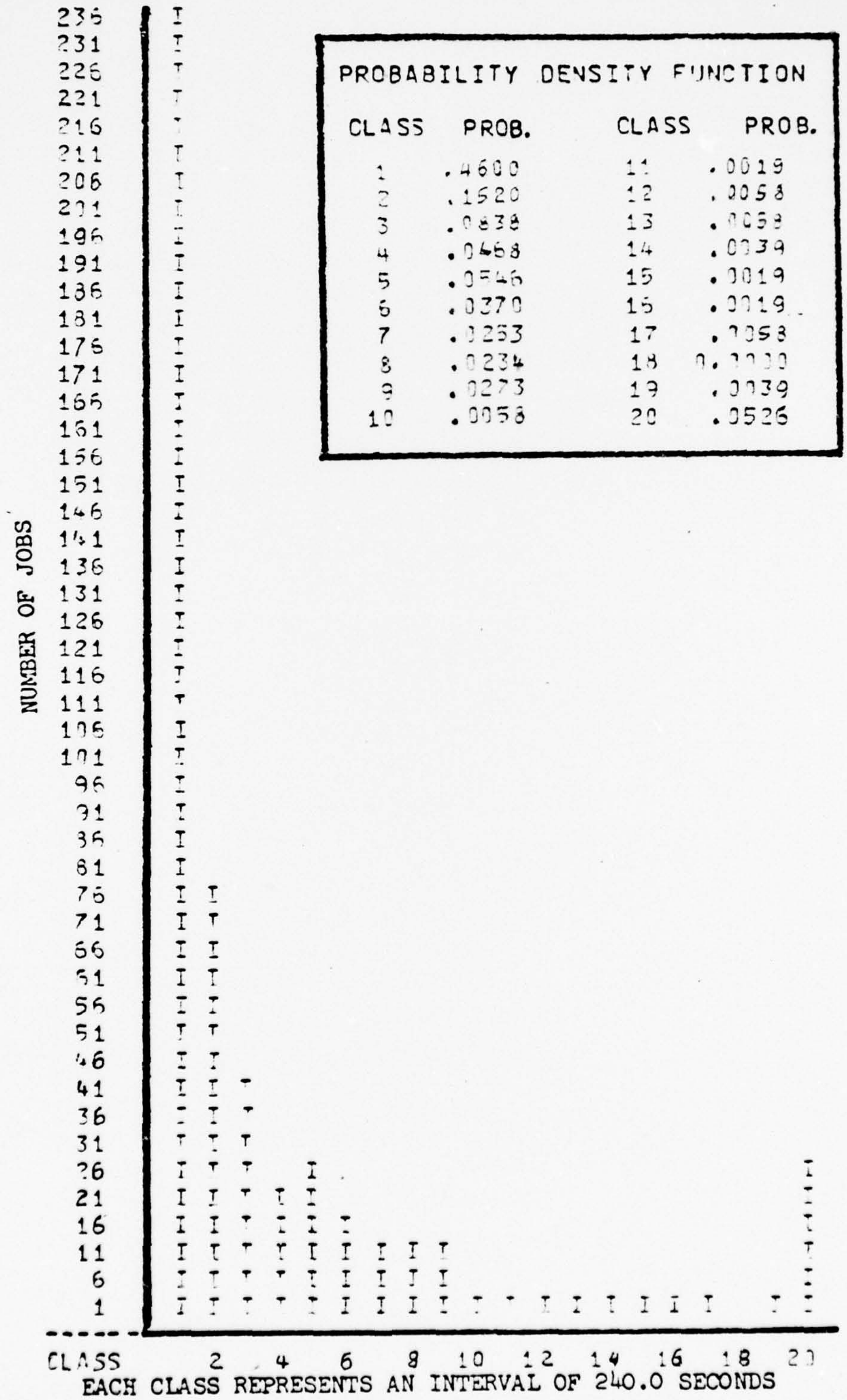
Fig. 35G. Distribution of Control Point Locations for 14 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.4444	11	.0136
2	.0994	12	.0136
3	.0804	13	.0097
4	.0526	14	.0097
5	.0546	15	.0039
6	.0467	16	.0039
7	.0312	17	.0039
8	.0468	18	.0058
9	.0195	19	.0039
10	.0097	20	.0643

Fig. 35H. Distribution of Interarrival time for 14 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.4600	11	.0019
2	.1520	12	.0058
3	.0838	13	.0058
4	.0468	14	.0039
5	.0546	15	.0019
6	.0370	16	.0019
7	.0253	17	.0058
8	.0234	18	0.0000
9	.0273	19	.0039
10	.0058	20	.0526

Fig. 35I. Distribution of Input Queue time for 14 October

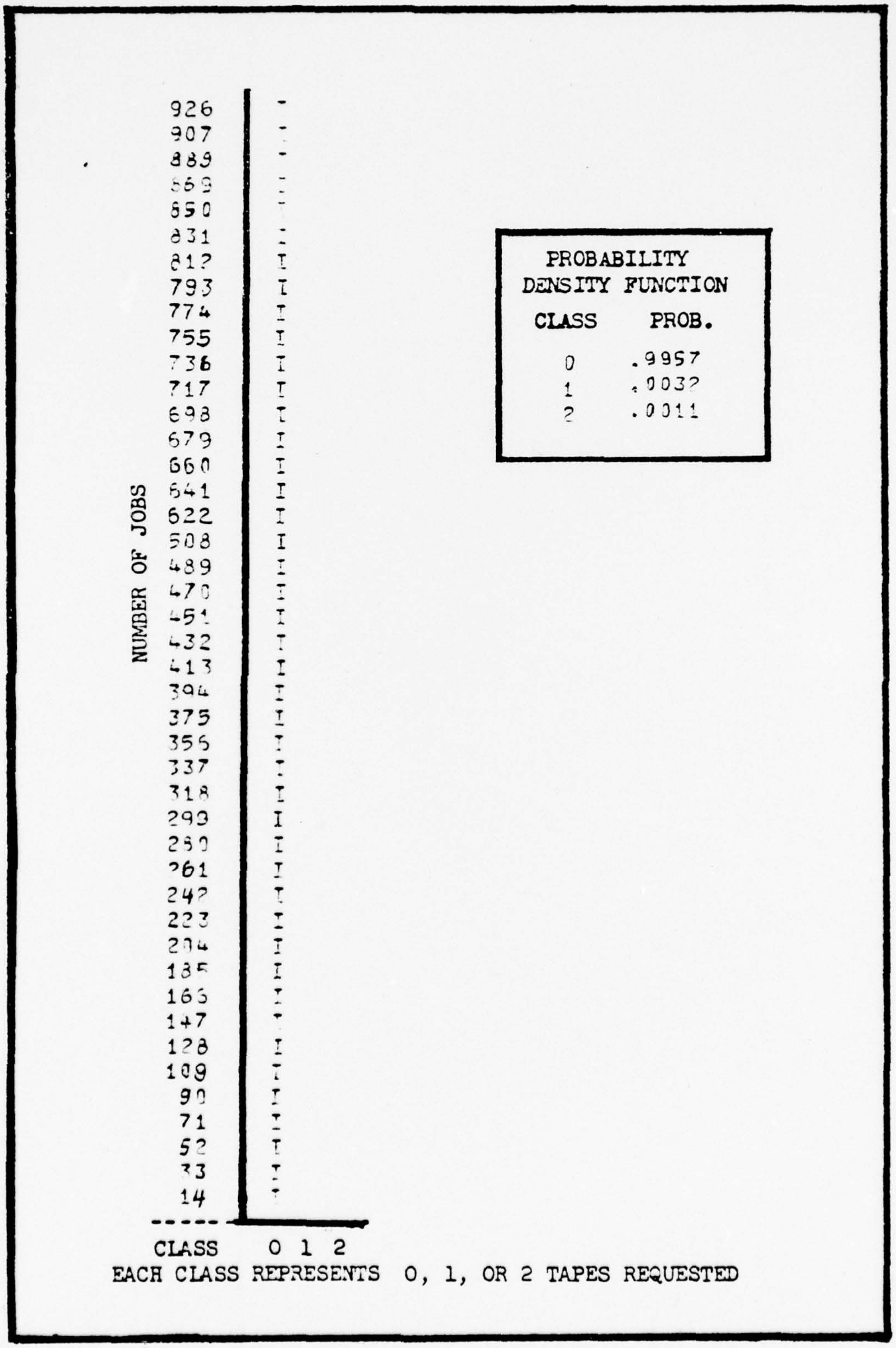


Fig. 35J. Distribution of Tapes Requested for 14 October

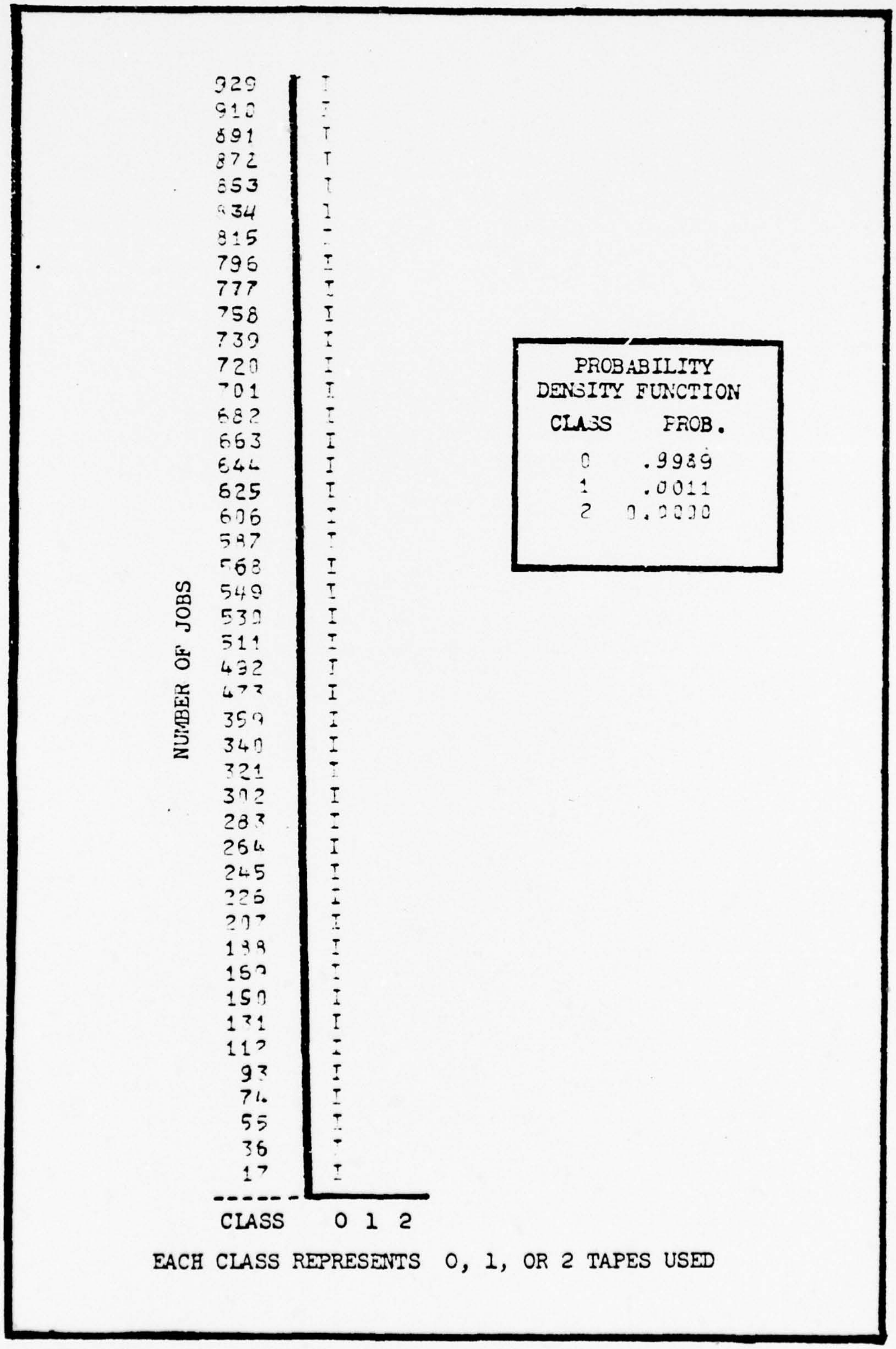


Fig. 35K. Distribution of Tapes Used for 14 October

TABLE XX

Statistical Summary for Workload Parameters,

14 October

Variable CPTIME CPU time in seconds

Mean	22.601	Std Err	2.178	Std Dev	65.412
Variance	4278.704	Kurtosis	90.374	Skewness	9.134
Minimum	.001	Maximum	1013.608	Sum	21154.721
C.V. Pct	289.417	.95 C.I.	18.405	to	26.797

Variable PPTIME Perpherial Processor time in seconds

Mean	62.359	Std Err	27.912	Std Dev	853.926
Variance	729192.755	Kurtosis	917.933	Skewness	30.229
Minimum	.169	Maximum	26067.820	Sum	58368.410
C.V. Pct	1369.365	.95 C.I.	7.583	to	117.136

Variable TIMEIO Input-Output time in seconds

Mean	15.490	Std Err	1.854	Std Dev	56.715
Variance	3216.639	Kurtosis	175.953	Skewness	11.746
Minimum	.037	Maximum	984.758	Sum	14498.670
C.V. Pct	365.141	.95 C.I.	11.852	to	19.128

Variable TOTCOST Total Cost in CRUs

Mean	24.955	Std Err	2.250	Std Dev	68.551
Variance	4740.465	Kurtosis	75.160	Skewness	7.560
Minimum	.020	Maximum	891.742	Sum	23357.481
C.V. Pct	275.905	.95 C.I.	20.538	to	29.371

TABLE XX (continued)

Variable KWS Memory in kilo-word seconds

Mean	701.934	Std Err	77.063	Std Dev	2357.687
Variance	5558689.011	Kurtosis	97.539	Skewness	8.779
Minimum	0.000	Maximum	34351.511	Sum	23357.481
C.V. Pct	275.905	.95 C.I.	20.538	to	29.371

Variable CPOT Control Point Occupancy time in seconds

Mean	777.241	Std Err	49.808	Std Dev	1523.841
Variance	2322092.975	Kurtosis	22.996	Skewness	4.123
Minimum	1.000	Maximum	13493.000	Sum	727498.000
C.V. Pct	196.058	.95 C.I.	679.492	to	874.990

Variable CMLOC Central Memory Locations in Use

Mean	[illegible]	Std Err	3083.887	Std Dev	94348.509
Variance	[illegible]	Kurtosis	-1.536	Skewness	.353
Minimum	[illegible]	Maximum	261120.000	Sum	.1160E+09
C.V. Pct	[illegible]	.95 C.I.	117977.757	to	130082.004

Variable CPLCC Control Points in Use

Mean	94.853	Std Err	.354	Std Dev	10.823
Variance	113.139	Kurtosis	32.948	Skewness	-5.253
Minimum	4.000	Maximum	100.000	Sum	88782.000
C.V. Pct	11.410	.95 C.I.	94.158	to	95.547

Variable ROLLOUT Total Time Rolled Out in seconds

Mean	17.366	Std Err	4.151	Std Dev	127.001
Variance	16129.313	Kurtosis	112.821	Skewness	10.157
Minimum	0.000	Maximum	1763.000	Sum	16255.000
C.V. Pct	731.302	.95 C.I.	9.220	to	25.913

TABLE XX (continued)

Variable IATIME Inter-arrival time in seconds

Mean	44.496	Std Err	3.147	Std Dev	73.795
Variance	5445.759	Kurtosis	9.221	Skewness	2.851
Minimum	0.000	Maximum	501.000	Sum	24474.000
C.V. Pct	165.839	.95 C.I.	38.317	to	50.679

Variable INQTIME Time in Input Queue in seconds

Mean	576.554	Std Err	55.316	Std Dev	1661.310
Variance	2759949.648	Kurtosis	25.719	Skewness	4.830
Minimum	0.000	Maximum	14523.000	Sum	520052.000
C.V. Pct	288.145	.95 C.I.	467.992	to	685.117

Variable TAPEREQ Number of Tapes Requested

Mean	.062	Std Err	.009	Std Dev	.267
Variance	.071	Kurtosis	29.854	Skewness	4.943
Minimum	0.000	Maximum	3.000	Sum	58.000
C.V. Pct	430.076	.95 C.I.	.045	to	.079

Variable TAPEUSED Number of Tapes Used

Mean	.058	Std Err	.008	Std Dev	.238
Variance	.057	Kurtosis	15.486	Skewness	4.031
Minimum	0.000	Maximum	2.000	Sum	54.000
C.V. Pct	412.231	.95 C.I.	.042	to	.073

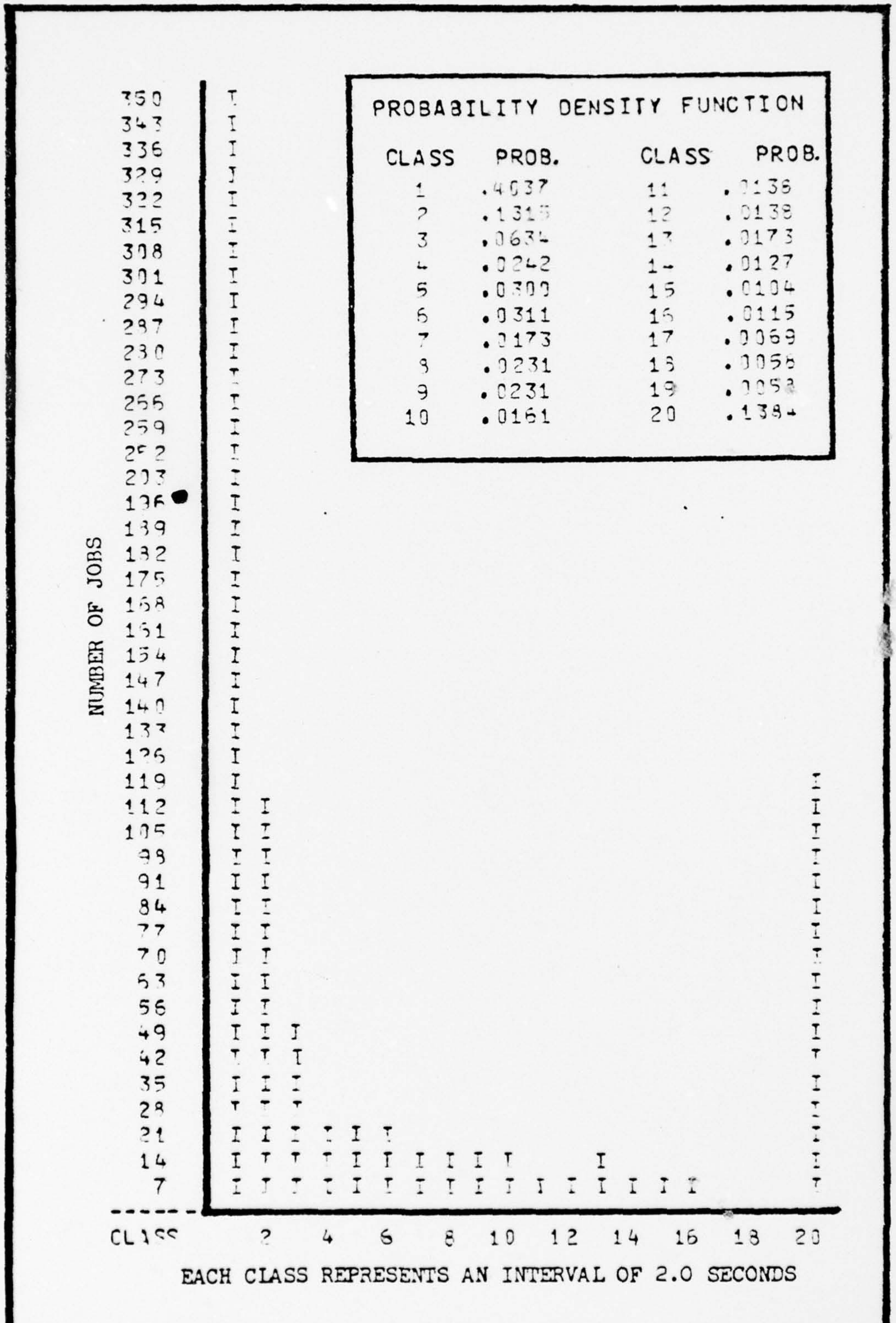
Appendix G

Workload Characterization

19 October

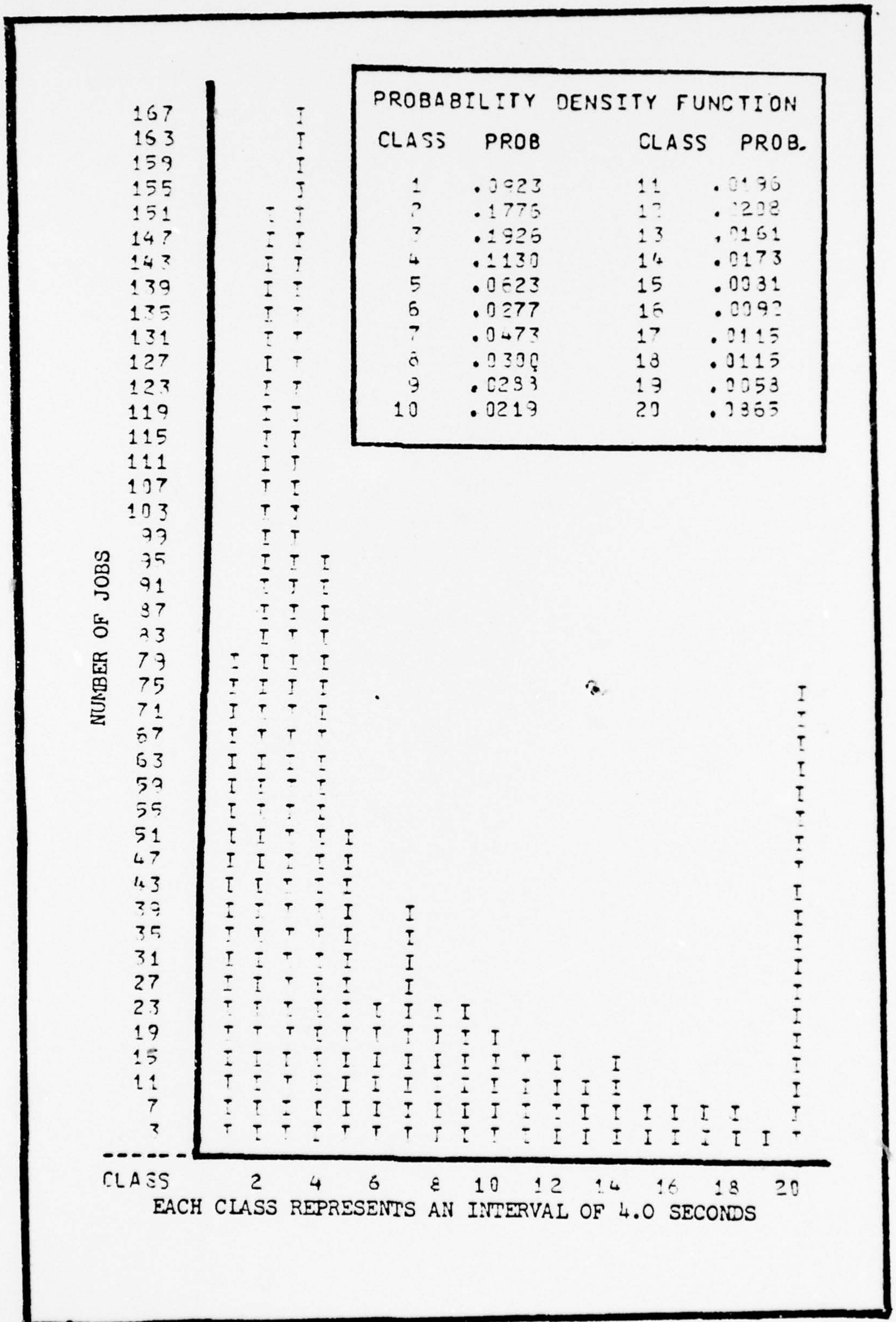
The following variables are characterized with a frequency histogram, probability density function, and/or a statistical summary. All measurements of time are in seconds.

Name	Description
CPTIME	The central processor time needed to process a job.
PPTIME	The peripheral processor time needed to process a job.
TIMEIO	The total I/O time needed to process a job.
TOTCOST	The total job cost in terms of CRUs.
KSW	The memory usage in terms of kilo-word seconds.
CPOT	The total control point occupancy time which is the time a job enters a control point until it leaves a control point for the last time.
CMLOC	The number of central memory locations occupied by all jobs including this job at the time this job left a control point.
CPLOC	The number of control points occupied by all jobs including this job at the time this job left a control point.
ROLLOUT	The total time that a batch job was rolled out for processing magnetic tapes.
IATIME	The total time between this batch job and the last batch job to arrive into the input queue.
INQTIME	The time a batch job spends in the input queue.
TAPEREQ/TAPEUSED	The number of tapes requested/used per job.



CLASS	PROB.	CLASS	PROB.
1	.4037	11	.0136
2	.1315	12	.0138
3	.0634	13	.0173
4	.0242	14	.0127
5	.0300	15	.0104
6	.0311	16	.0115
7	.0173	17	.0069
8	.0231	18	.0056
9	.0231	19	.0058
10	.0161	20	.1384

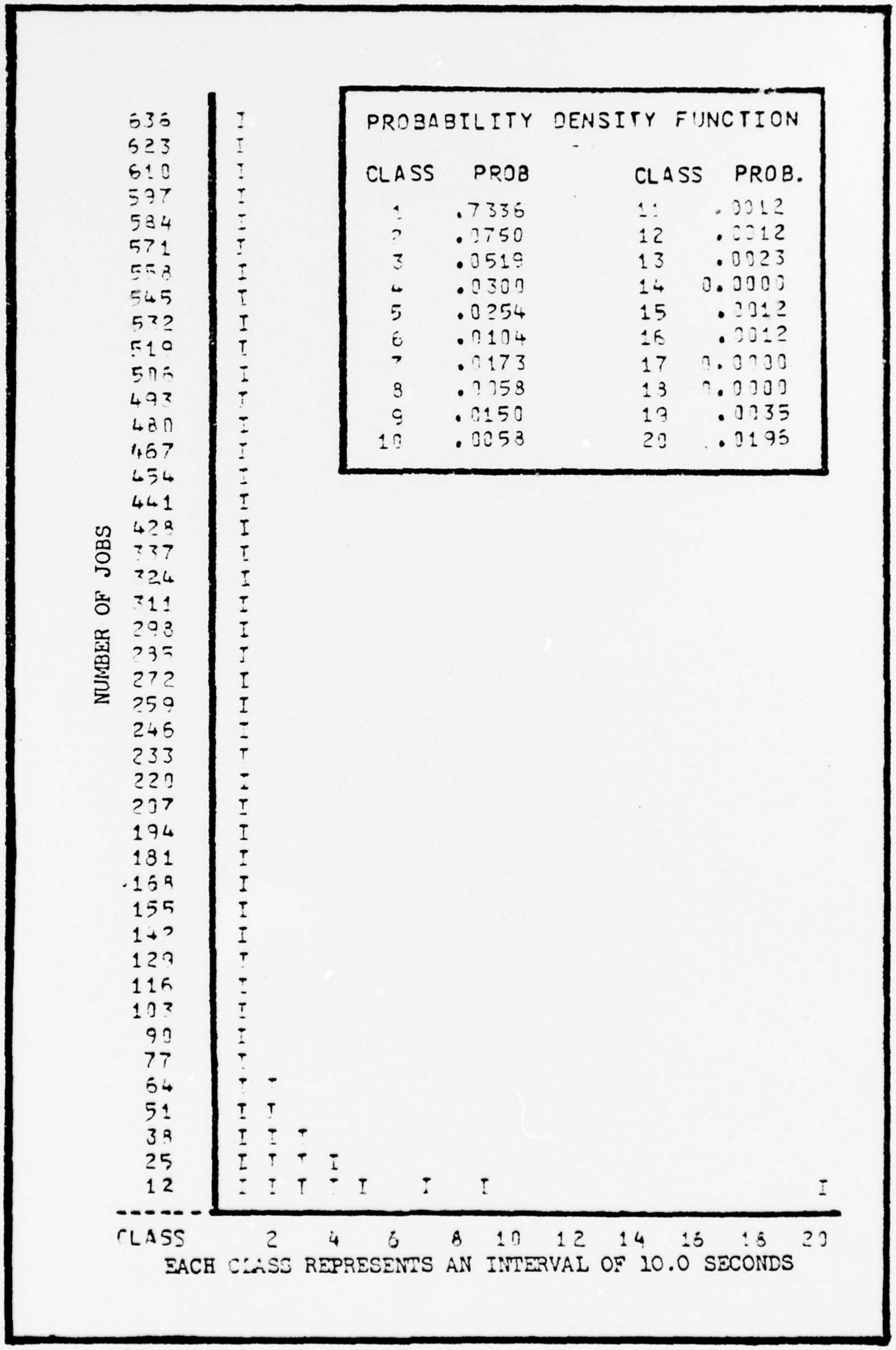
Fig. 36A. Distribution of CPU time for 19 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.0923	11	.0196
2	.1776	12	.0208
3	.1926	13	.0161
4	.1130	14	.0173
5	.0623	15	.0081
6	.0277	16	.0092
7	.0473	17	.0115
8	.0300	18	.0115
9	.0283	19	.0058
10	.0219	20	.0865

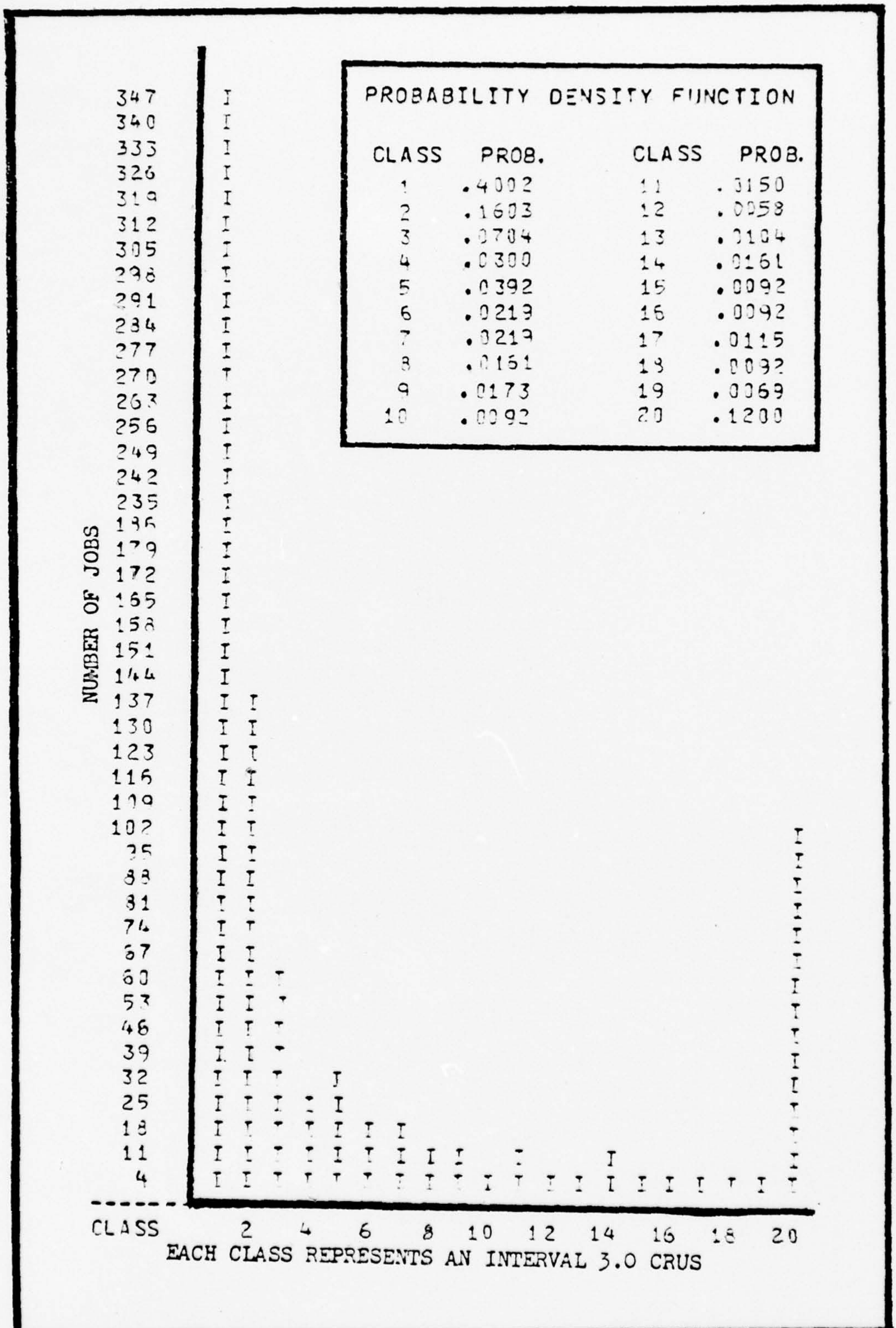
Fig. 36B. Distribution of PPU time for 19 October



PROBABILITY DENSITY FUNCTION

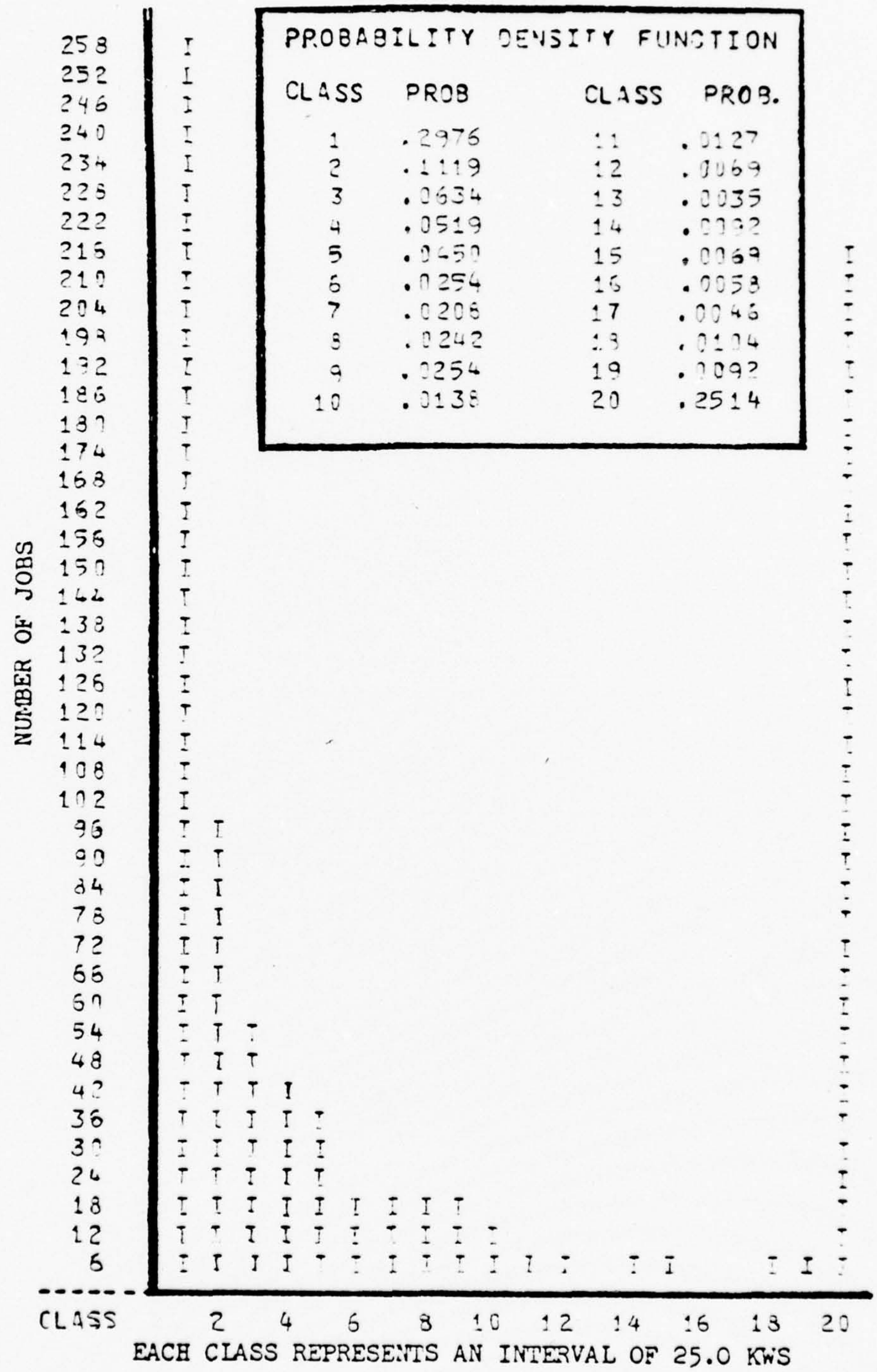
CLASS	PROB	CLASS	PROB.
1	.7336	11	.0012
2	.0750	12	.0012
3	.0519	13	.0023
4	.0300	14	0.0000
5	.0254	15	.0012
6	.0104	16	.0012
7	.0173	17	0.0000
8	.0058	18	0.0000
9	.0150	19	.0035
10	.0058	20	.0196

Fig. 36C. Distribution of I/O time for 19 October



CLASS	PROB.	CLASS	PROB.
1	.4002	11	.0150
2	.1603	12	.0058
3	.0704	13	.0104
4	.0300	14	.0161
5	.0392	15	.0092
6	.0219	16	.0092
7	.0219	17	.0115
8	.0161	18	.0092
9	.0173	19	.0069
10	.0092	20	.1200

Fig. 36D. Distribution of CRUs for 19 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.2976	11	.0127
2	.1119	12	.0069
3	.0634	13	.0035
4	.0519	14	.0092
5	.0450	15	.0069
6	.0254	16	.0058
7	.0206	17	.0046
8	.0242	18	.0104
9	.0254	19	.0092
10	.0138	20	.2514

Fig. 36E. Distribution of KWS for 19 October

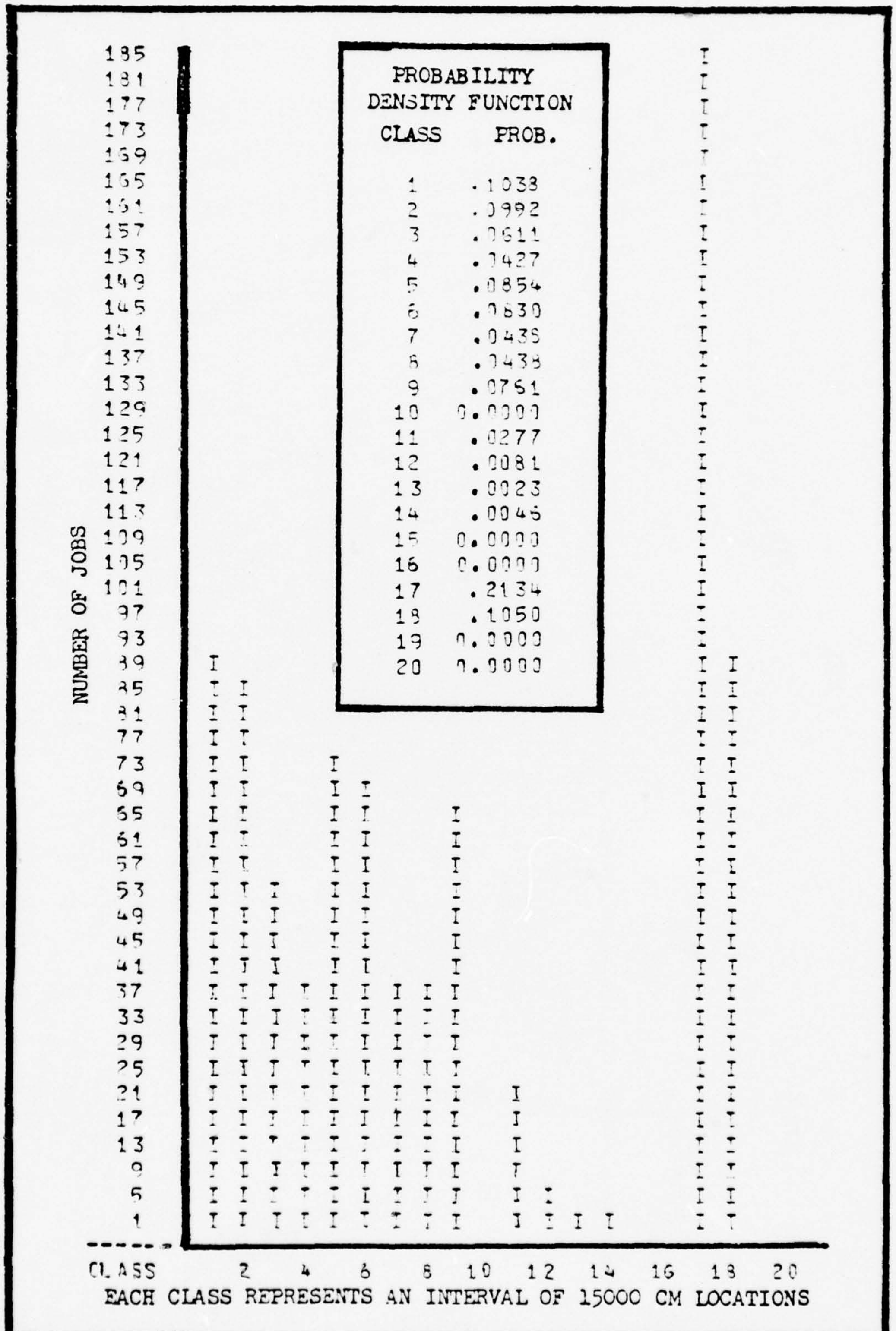


Fig. 36F. Distribution of Central Memory Locations for 19 October

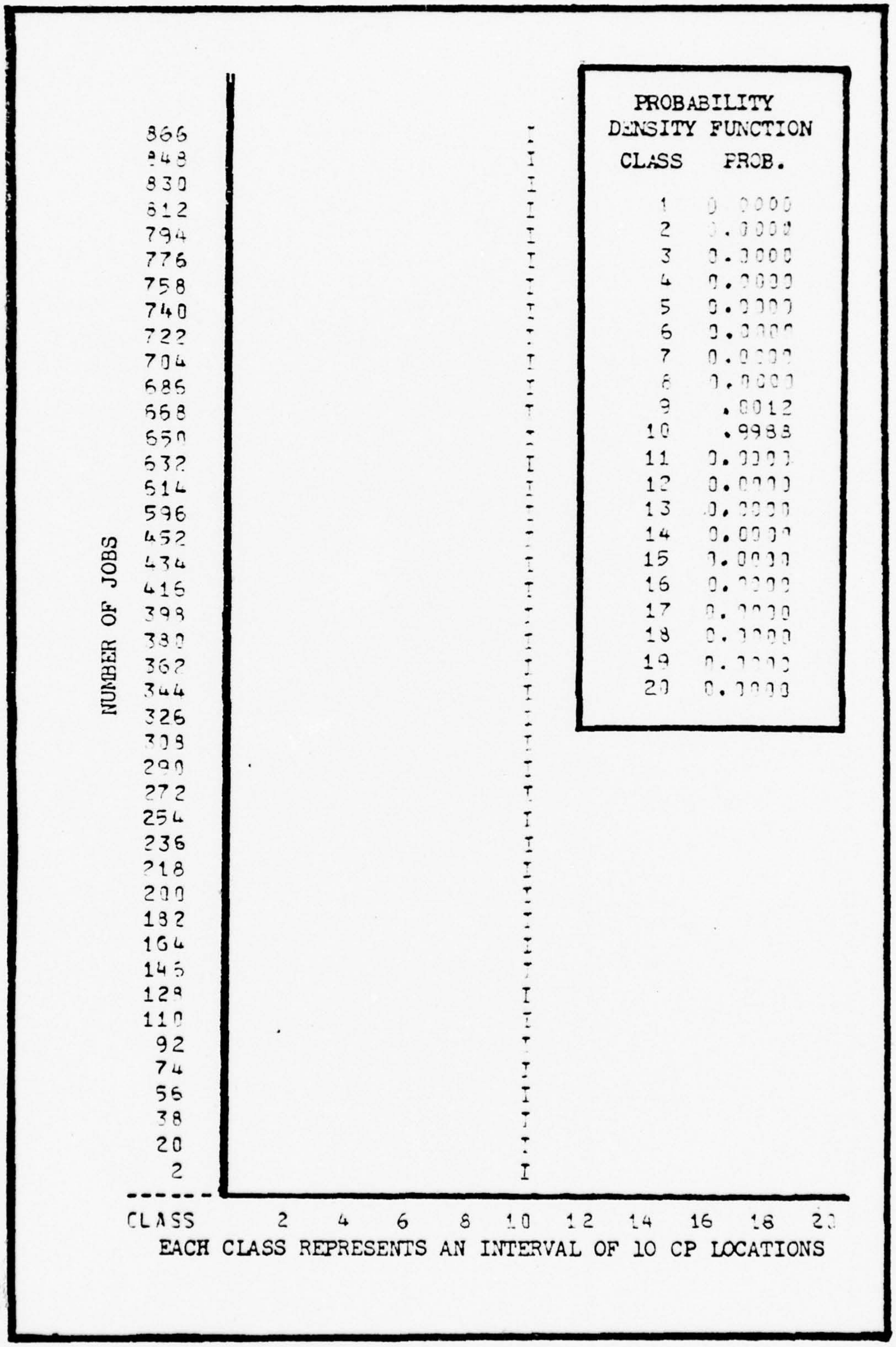
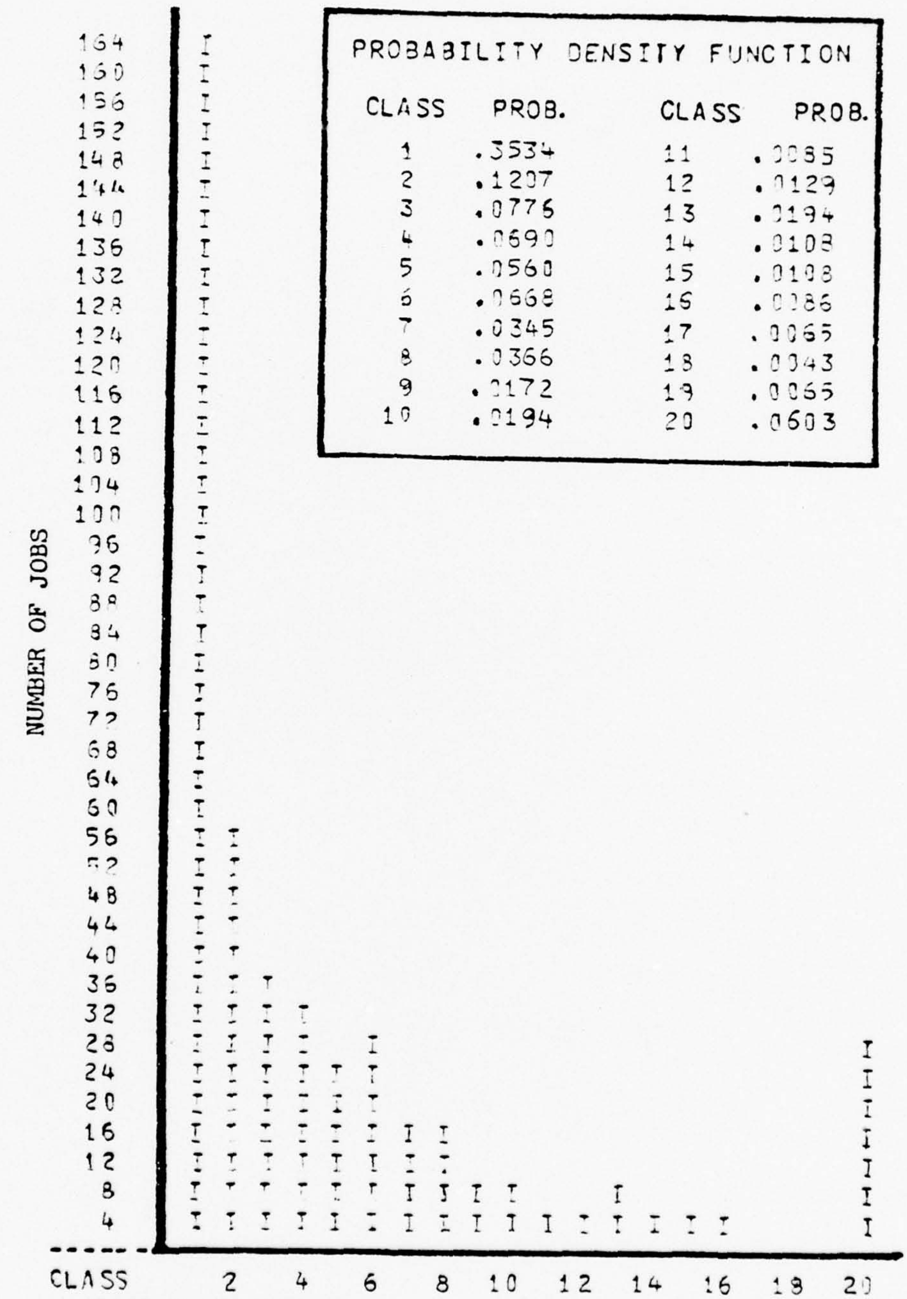


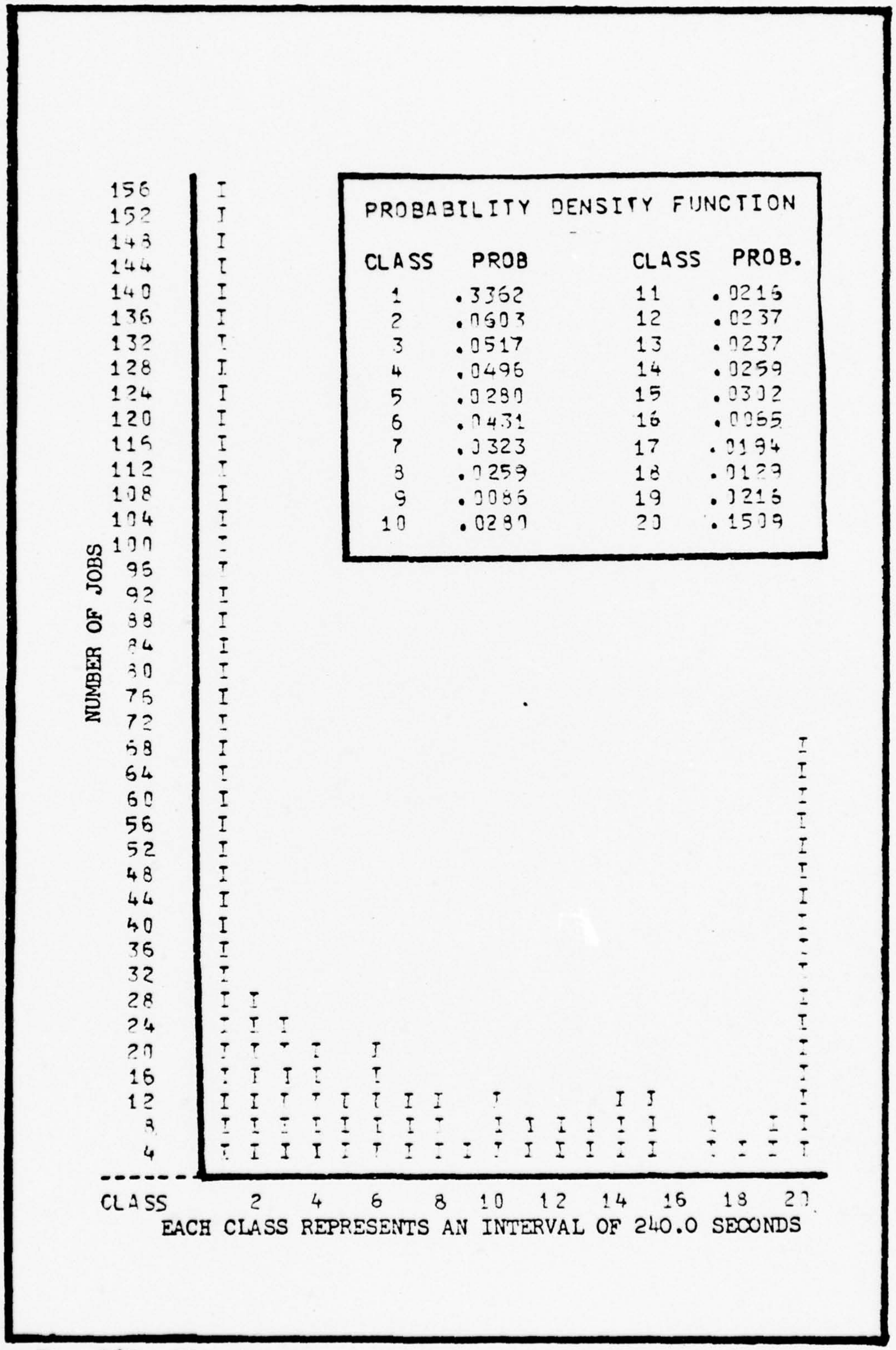
Fig. 36G. Distribution of Control Points Locations for 19 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.3534	11	.0085
2	.1207	12	.0129
3	.0776	13	.0194
4	.0690	14	.0108
5	.0560	15	.0108
6	.0668	16	.0086
7	.0345	17	.0065
8	.0366	18	.0043
9	.0172	19	.0065
10	.0194	20	.0603

Fig. 36H. Distribution of Interarrival time for 19 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.3362	11	.0216
2	.0603	12	.0237
3	.0517	13	.0237
4	.0496	14	.0259
5	.0280	15	.0302
6	.0431	16	.0065
7	.0323	17	.0194
8	.0259	18	.0129
9	.0086	19	.0216
10	.0280	20	.1509

Fig. 36I. Distribution of Input Queue time for 19 October

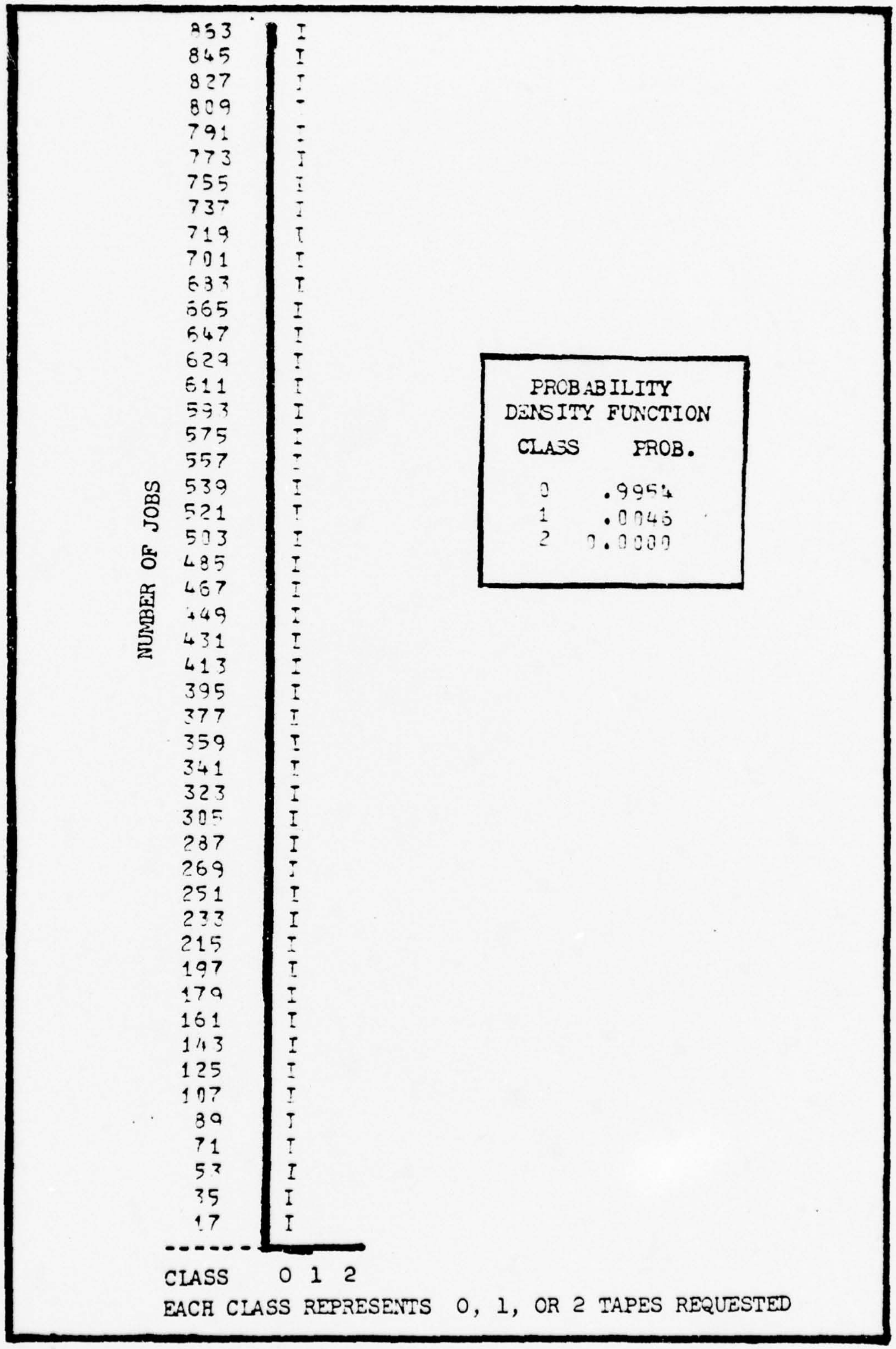


Fig. 36J. Distribution of Tapes Requested for 19 October

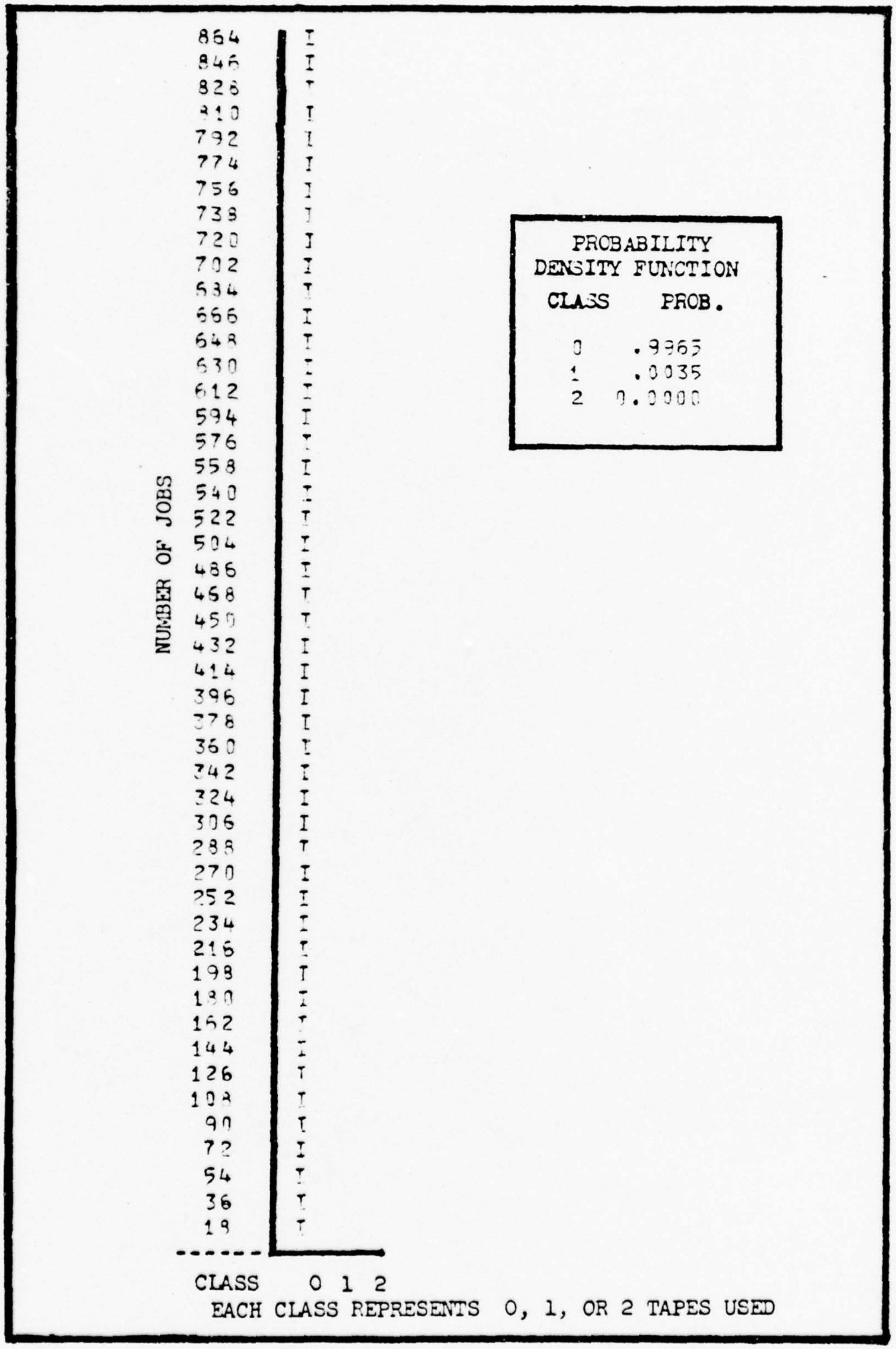


Fig. 36K. Distribution of Tapes Used for 19 October

TABLE XXI

Statistical Summary for Workload Parameters,

19 October

Variable CPTIME CPU time in seconds

Mean	23.123	Std Err	2.285	Std Dev	68.501
Variance	4692.422	Kurtosis	102.481	Skewness	8.564
Minimum	.001	Maximum	1015.043	Sum	20788.016
C.V. Pct	296.241	.95 C.I.	18.640	to	27.607

Variable PPTIME Peripheral

Mean	34.530	Std Err	2.938	Std Dev	88.101
Variance	7761.847	Kurtosis	237.969	Skewness	12.884
Minimum	.491	Maximum	1914.830	Sum	31042.340
C.V. Pct	255.146	.95 C.I.	28.763	to	40.297

Variable TIMEIO Input-Output time in seconds

Mean	22.947	Std Err	3.319	Std Dev	99.613
Variance	9902.743	Kurtosis	128.887	Skewness	10.508
Minimum	.074	Maximum	1465.027	Sum	20629.608
C.V. Pct	433.657	.95 C.I.	16.434	to	29.461

Variable TOTCOST Total Cost in CRUs

Mean	31.428	Std Err	3.468	Std Dev	103.177
Variance	10811.200	Kurtosis	78.412	Skewness	8.080
Minimum	.040	Maximum	1306.023	Sum	28253.911
C.V. Pct	330.840	.95 C.I.	24.622	to	38.234

TABLE XXI (continued)

Variable KWS Memory in kilo-word seconds

Mean	888.875	Std Err	110.824	Std Dev	3322.858
Variance	1104E+08	Kurtosis	72.515	Skewness	8.015
Minimum	0.000	Maximum	36981.184	Sum	799098.944
C.V. Pct	373.827	.95 C.I.	671.372	to	1106.379

Variable CPOT Control Point Occupancy time in seconds

Mean	863.497	Std Err	54.119	Std Dev	1622.663
Variance	2633036.805	Kurtosis	17.188	Skewness	3.573
Minimum	1.000	Maximum	15725.000	Sum	776284.000
C.V. Pct	187.919	.95 C.I.	757.283	to	969.711

Variable CMLOC Central Memory Locations in Use

Mean	129791.858	Std Err	3176.752	Std Dev	95249.887
Variance	9072E+10	Kurtosis	-1.556	Skewness	.249
Minimum	1024.000	Maximum	256000.000	Sum	.1166E+09
C.V. Pct	73.387	.95 C.I.	113557.116	to	136026.599

Variable CPLOC Control Points in Use

Mean	97.250	Std Err	.214	Std Dev	6.426
Variance	41.290	Kurtosis	168.674	Skewness	-12.260
Minimum	7.000	Maximum	100.000	Sum	87428.000
C.V. Pct	6.607	.95 C.I.	96.830	to	97.671

Variable ROLLOUT Total Time Rolled Out in seconds

Mean	27.496	Std Err	12.551	Std Dev	376.322
Variance	141617.918	Kurtosis	409.655	Skewness	19.762
Minimum	0.000	Maximum	8504.000	Sum	24719.000
C.V. Pct	1368.636	.95 C.I.	2.863	to	52.129

TABLE XXI (continued)

Variable IATIME Inter-arrival time in seconds

Mean	49.523	Std Err	3.615	Std Dev	79.701
Variance	6352.271	Kurtosis	29.314	Skewness	4.290
Minimum	0.000	Maximum	843.000	Sum	24068.000
C.V. Pct	160.939	.95 C.I.	42.419	to	56.625

Variable INQTIME Time in Input Queue in seconds

Mean	1710.987	Std Err	99.553	Std Dev	2888.761
Variance	8344941.328	Kurtosis	16.611	Skewness	3.675
Minimum	0.000	Maximum	23103.000	Sum	1103851.000
C.V. Pct	220.350	.95 C.I.	1115.585	to	1506.389

Variable TAPEREQ Number of Tapes Requested

Mean	.048	Std Err	.008	Std Dev	.233
Variance	.055	Kurtosis	29.601	Skewness	5.243
Mini	0.000	Maximum	2.000	Sum	43.000
C.V. Pct	488.090	.95 C.I.	.033	to	.063

Variable TAPEUSED Number of Tapes Used

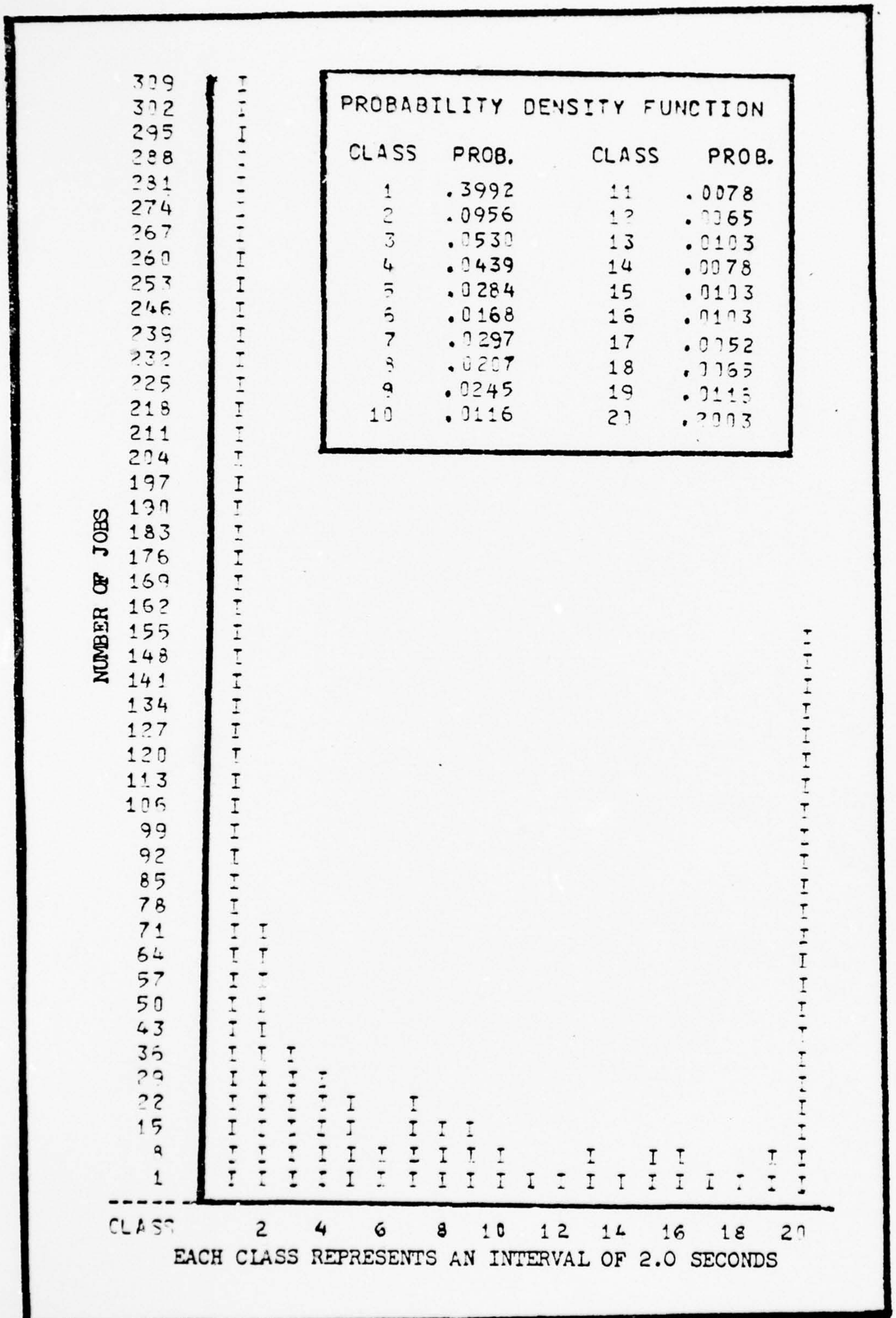
Mean	.047	Std Err	.008	Std Dev	.226
Variance	.051	Kurtosis	28.135	Skewness	5.131
Minimum	0.000	Maximum	2.000	Sum	42.000
C.V. Pct	484.652	.95 C.I.	.032	to	.062

Appendix H

Workload Characterization, 13 October

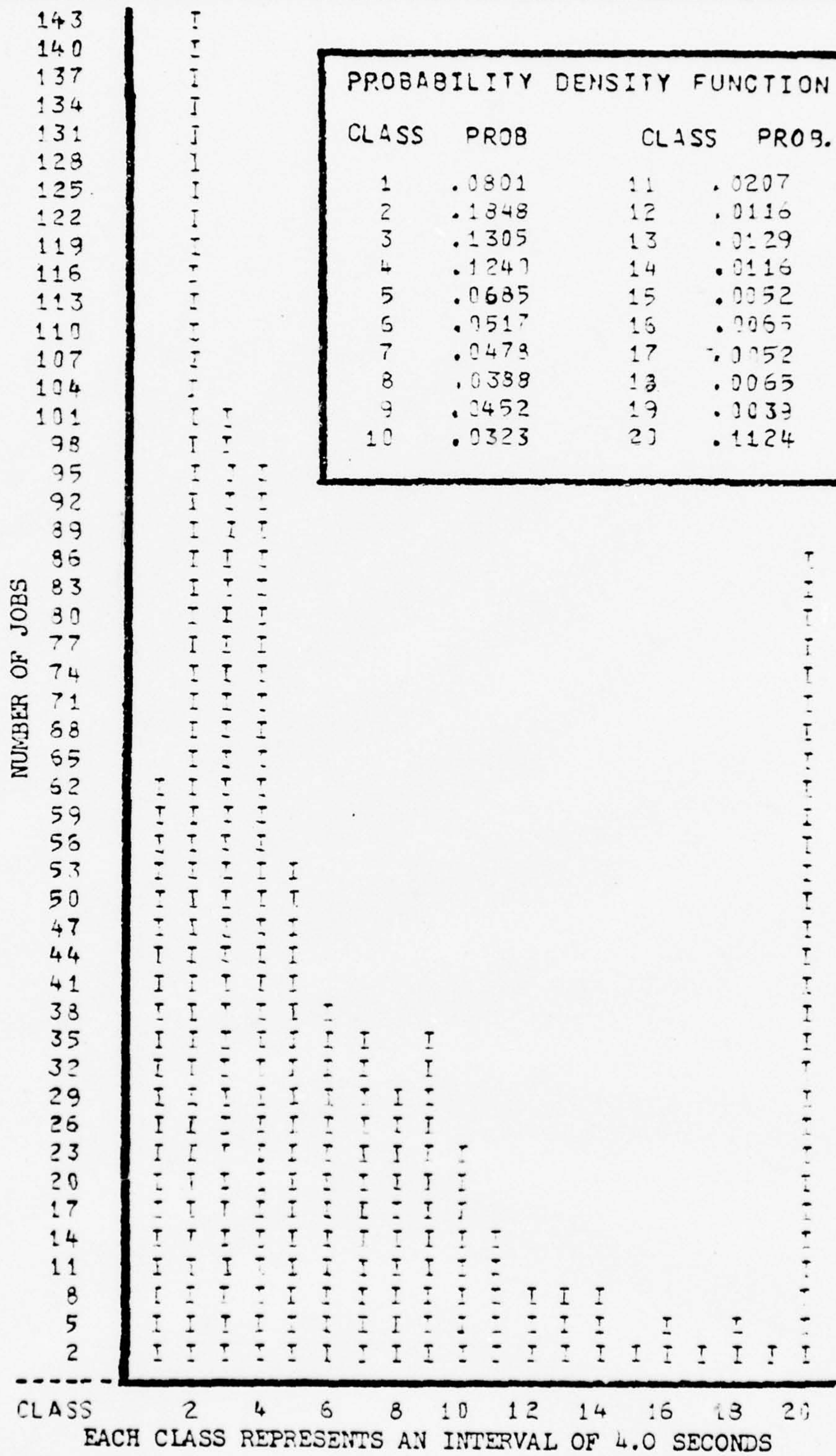
The following variables are characterized with a frequency histogram, probability density function, and/or a statistical summary. All measurements of time are in seconds.

Name	Description
CPTIME	The central processor time needed to process a job.
PPTIME	The peripheral processor time needed to process a job.
TIMEIO	The total I/O time needed to process a job.
TOTCOST	The total job cost in terms of CRUs.
KSW	The memory usage in terms of kilo-word seconds.
CPOT	The total control point occupancy time which is the time a job enters a control point until it leaves a control point for the last time.
CMLOC	The number of central memory locations occupied by all jobs including this job at the time this job left a control point.
CPLOC	The number of control points occupied by all jobs including this job at the time this job left a control point.
ROLLOUT	The total time that a batch job was rolled out for processing magnetic tapes.
IATIME	The total time between this batch job and the last batch job to arrive into the input queue.
INQTIME	The time a batch job spends in the input queue.
TAPEREQ/TAPEUSED	The number of tapes requested/used per job.



CLASS	PROB.	CLASS	PROB.
1	.3992	11	.0078
2	.0956	12	.0065
3	.0530	13	.0103
4	.0439	14	.0078
5	.0284	15	.0103
6	.0168	16	.0103
7	.0297	17	.0052
8	.0207	18	.0065
9	.0245	19	.0116
10	.0116	20	.2003

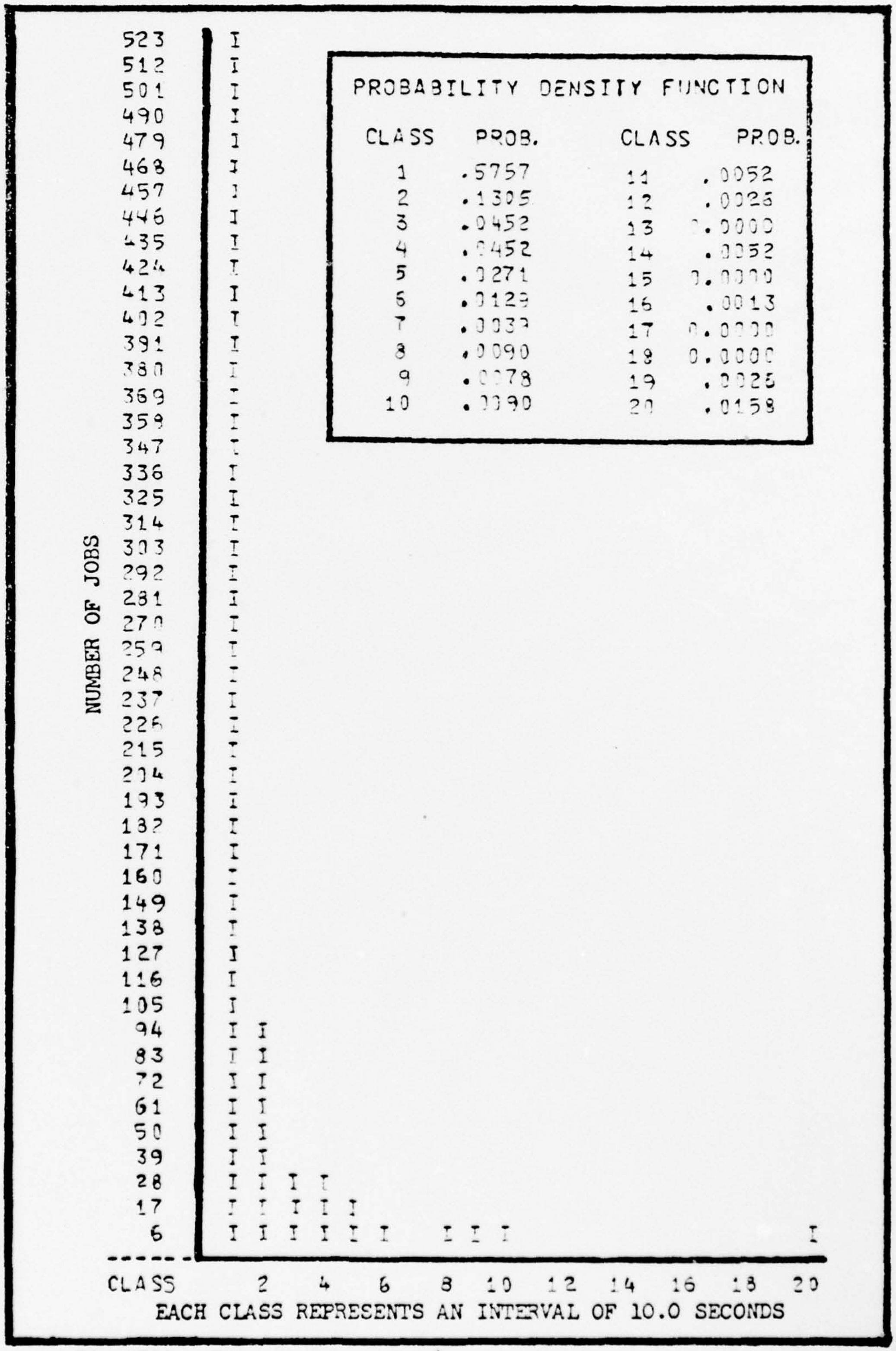
Fig. 37A. Distribution of CPU time for 13 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.0801	11	.0207
2	.1848	12	.0116
3	.1305	13	.0129
4	.1240	14	.0116
5	.0685	15	.0052
6	.0517	16	.0065
7	.0478	17	.0052
8	.0388	18	.0065
9	.0452	19	.0039
10	.0323	20	.1124

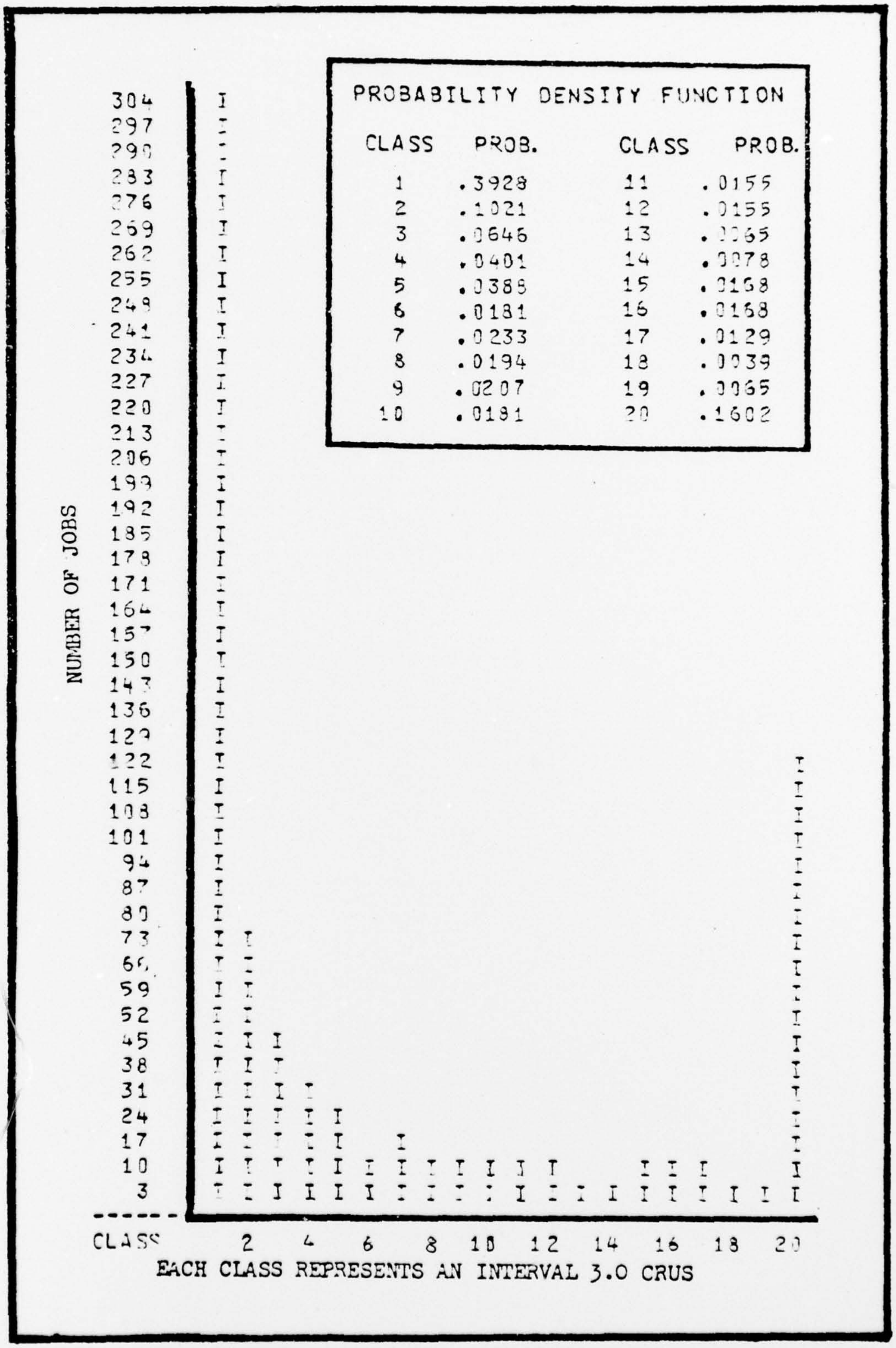
Fig. 37B. Distribution of PPU time for 13 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.5757	11	.0052
2	.1305	12	.0026
3	.0452	13	0.0000
4	.0452	14	.0052
5	.0271	15	0.0000
6	.0129	16	.0013
7	.0039	17	0.0000
8	.0090	18	0.0000
9	.0078	19	.0026
10	.0090	20	.0158

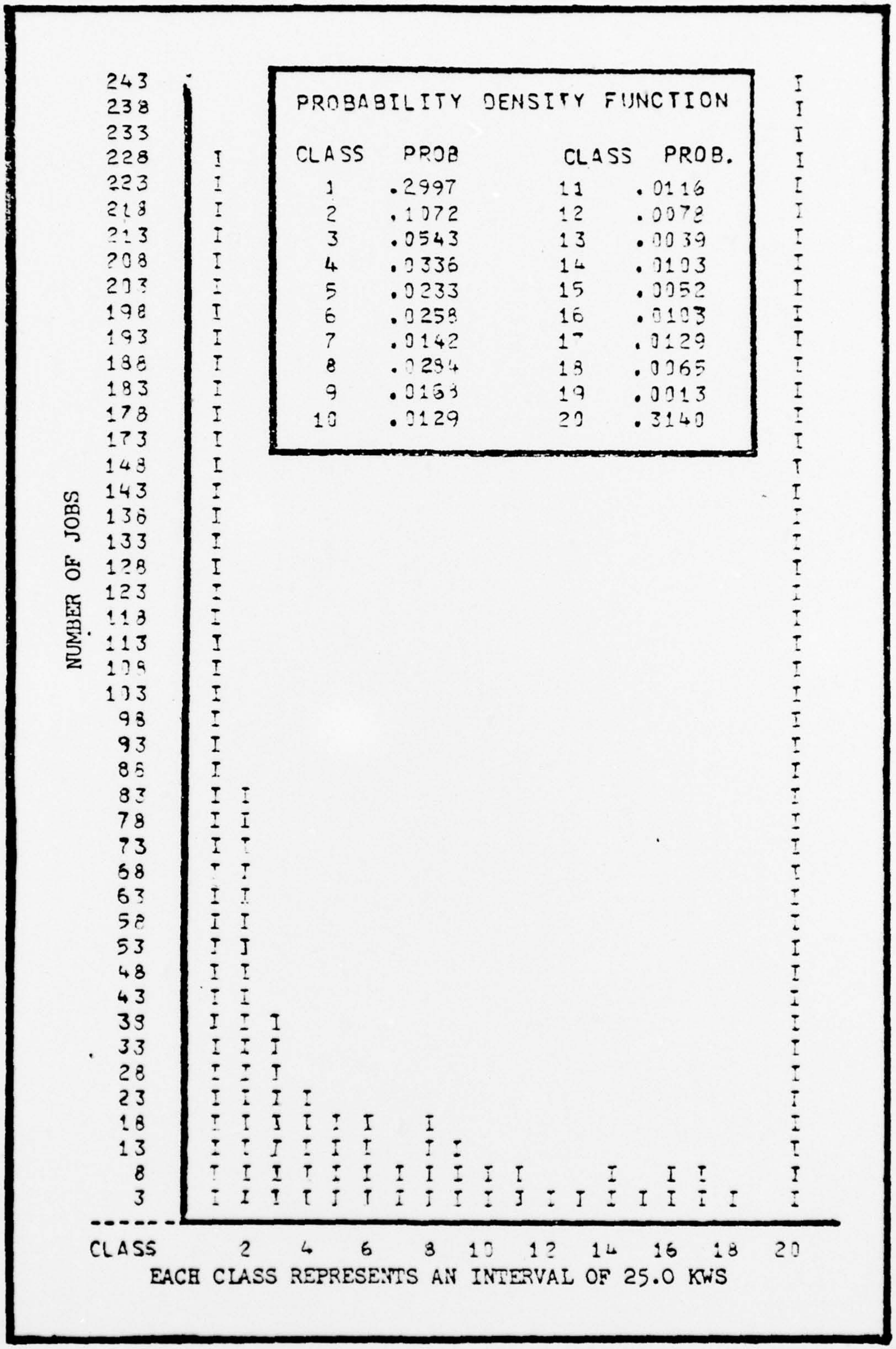
Fig. 37C. Distribution of I/O time for 13 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.3928	11	.0155
2	.1021	12	.0155
3	.0646	13	.0065
4	.0401	14	.0078
5	.0388	15	.0168
6	.0181	16	.0168
7	.0233	17	.0129
8	.0194	18	.0039
9	.0207	19	.0065
10	.0181	20	.1602

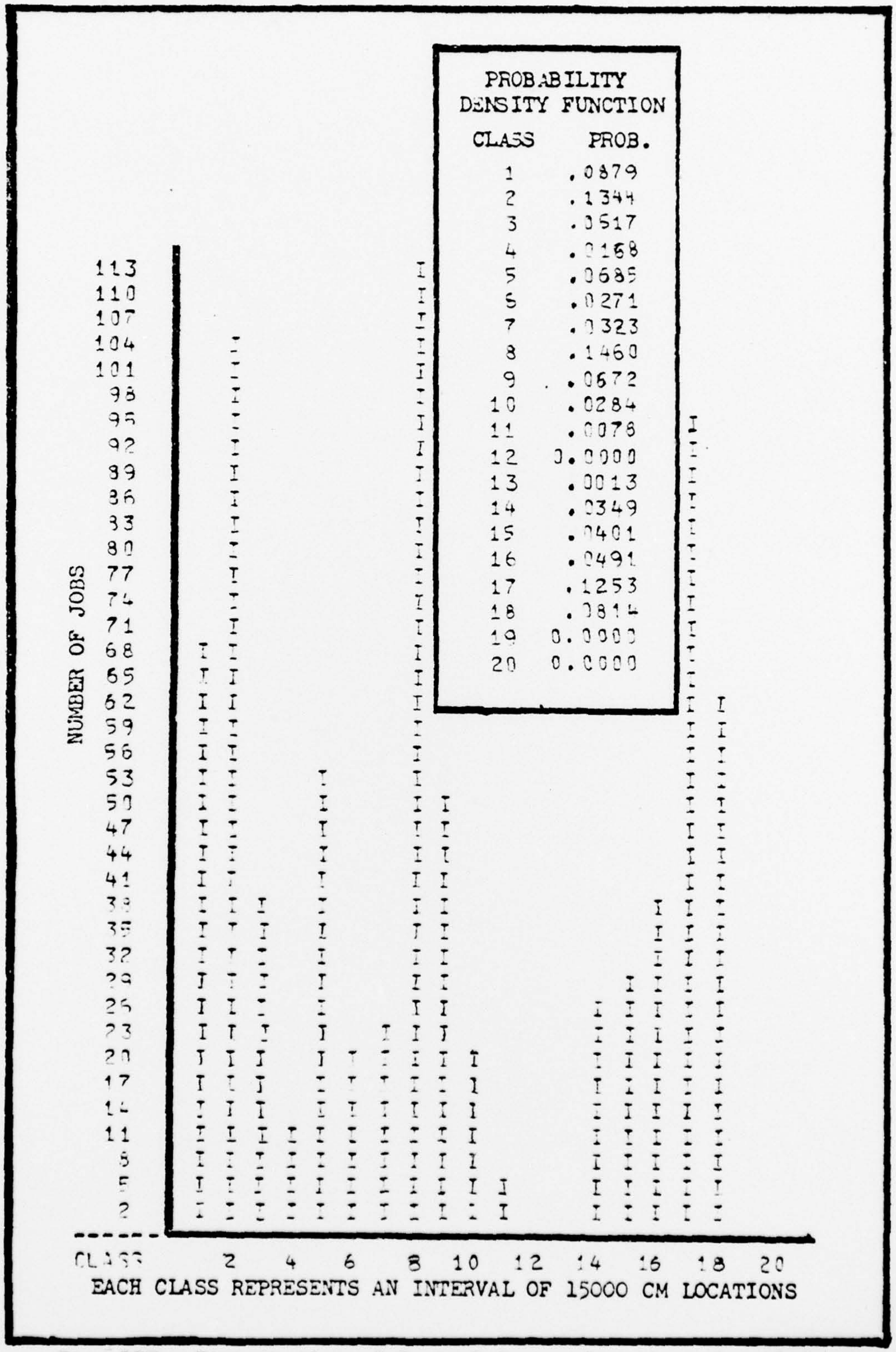
Fig. 37D. Distribution of CRUs for 13 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.2997	11	.0116
2	.1072	12	.0078
3	.0543	13	.0039
4	.0336	14	.0103
5	.0233	15	.0052
6	.0258	16	.0103
7	.0142	17	.0129
8	.0284	18	.0065
9	.0168	19	.0013
10	.0129	20	.3140

Fig. 37E. Distribution of KWS for 13 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.
1	.0879
2	.1344
3	.0517
4	.0168
5	.0685
6	.0271
7	.0323
8	.1460
9	.0672
10	.0284
11	.0078
12	0.0000
13	.0013
14	.0349
15	.0401
16	.0491
17	.1253
18	.0814
19	0.0000
20	0.0000

Fig. 37F. Distribution of Central Memory Locations for 13 October

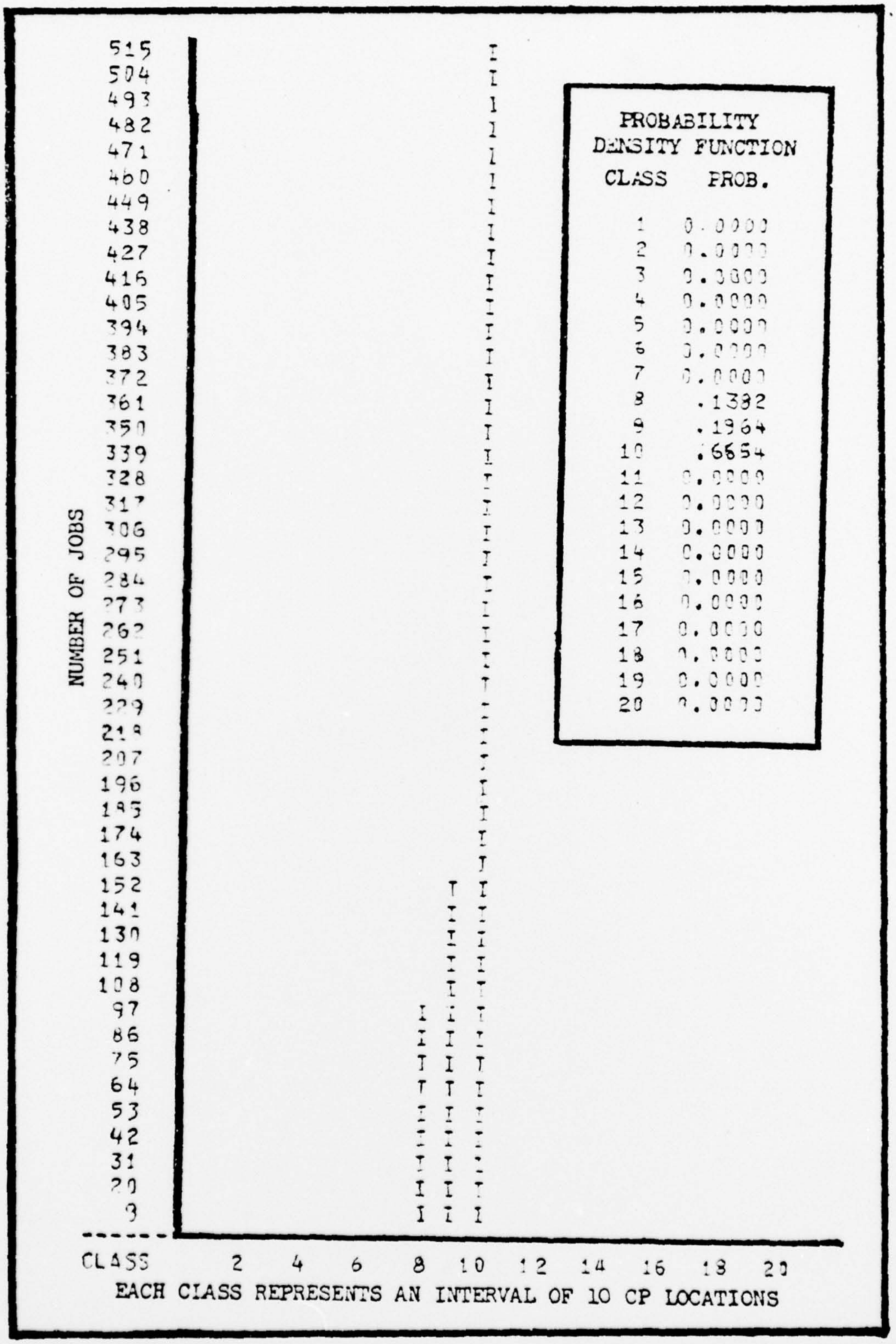
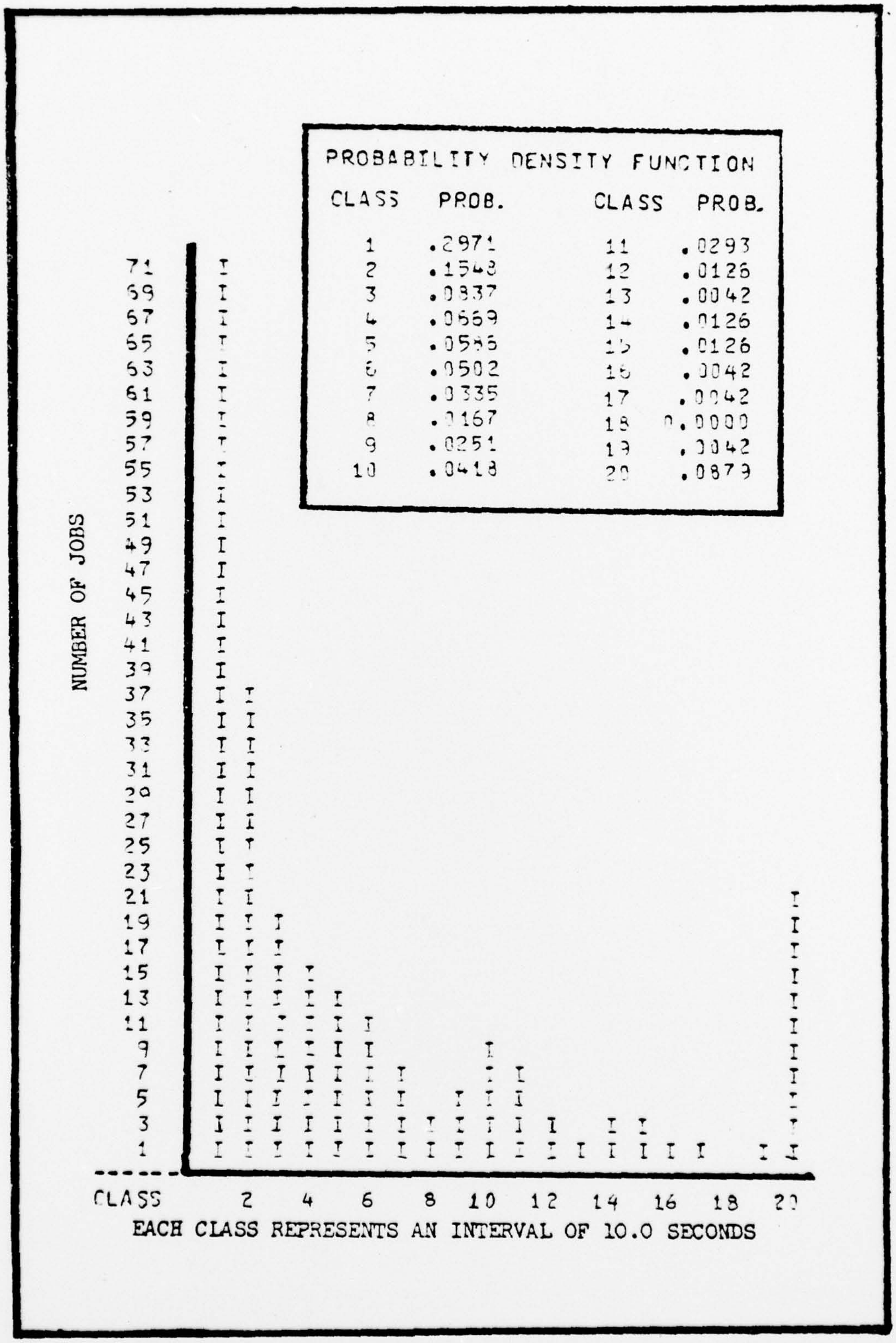


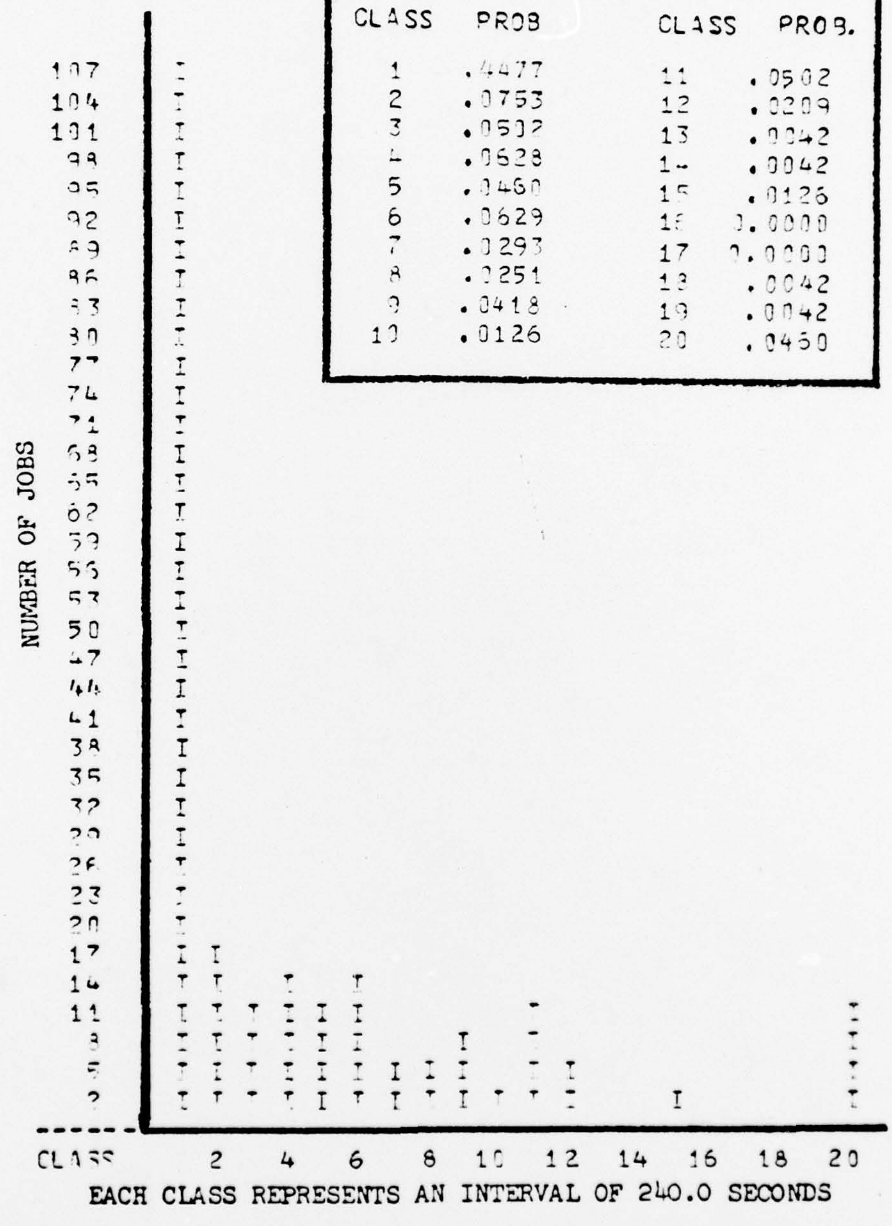
Fig. 37G. Distribution of Control Point Locations for 13 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB.	CLASS	PROB.
1	.2971	11	.0293
2	.1548	12	.0126
3	.0837	13	.0042
4	.0669	14	.0126
5	.0586	15	.0126
6	.0502	16	.0042
7	.0335	17	.0042
8	.0167	18	0.0000
9	.0251	19	.0042
10	.0418	20	.0879

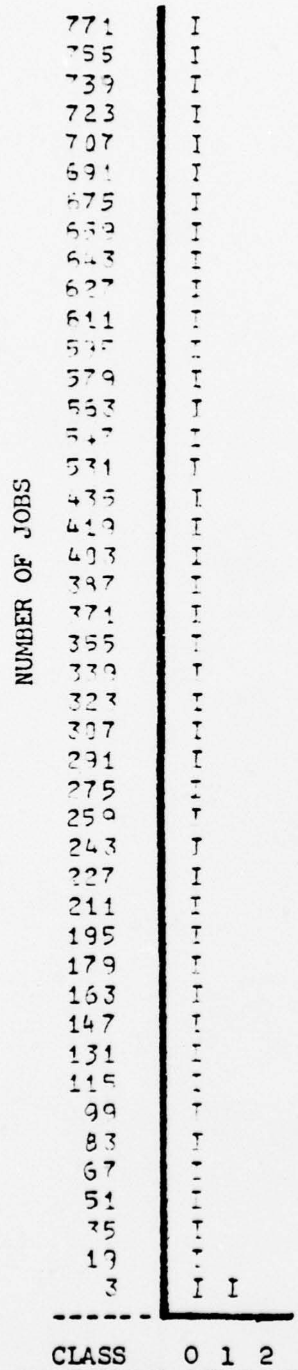
Fig. 37H. Distribution of Interarrival time for 13 October



PROBABILITY DENSITY FUNCTION

CLASS	PROB	CLASS	PROB.
1	.4477	11	.0502
2	.0753	12	.0209
3	.0502	13	.0042
4	.0628	14	.0042
5	.0460	15	.0126
6	.0629	16	0.0000
7	.0293	17	0.0000
8	.0251	18	.0042
9	.0418	19	.0042
10	.0126	20	.0460

Fig. 37I. Distribution of Input Queue time for 13 October



PROBABILITY DENSITY FUNCTION CLASS	PROB.
0	.9961
1	.0039
2	0.0000

EACH CLASS REPRESENTS 0, 1, OR 2 TAPES REQUESTED

Fig. 37J. Distribution of Tapes Requested for 13 October

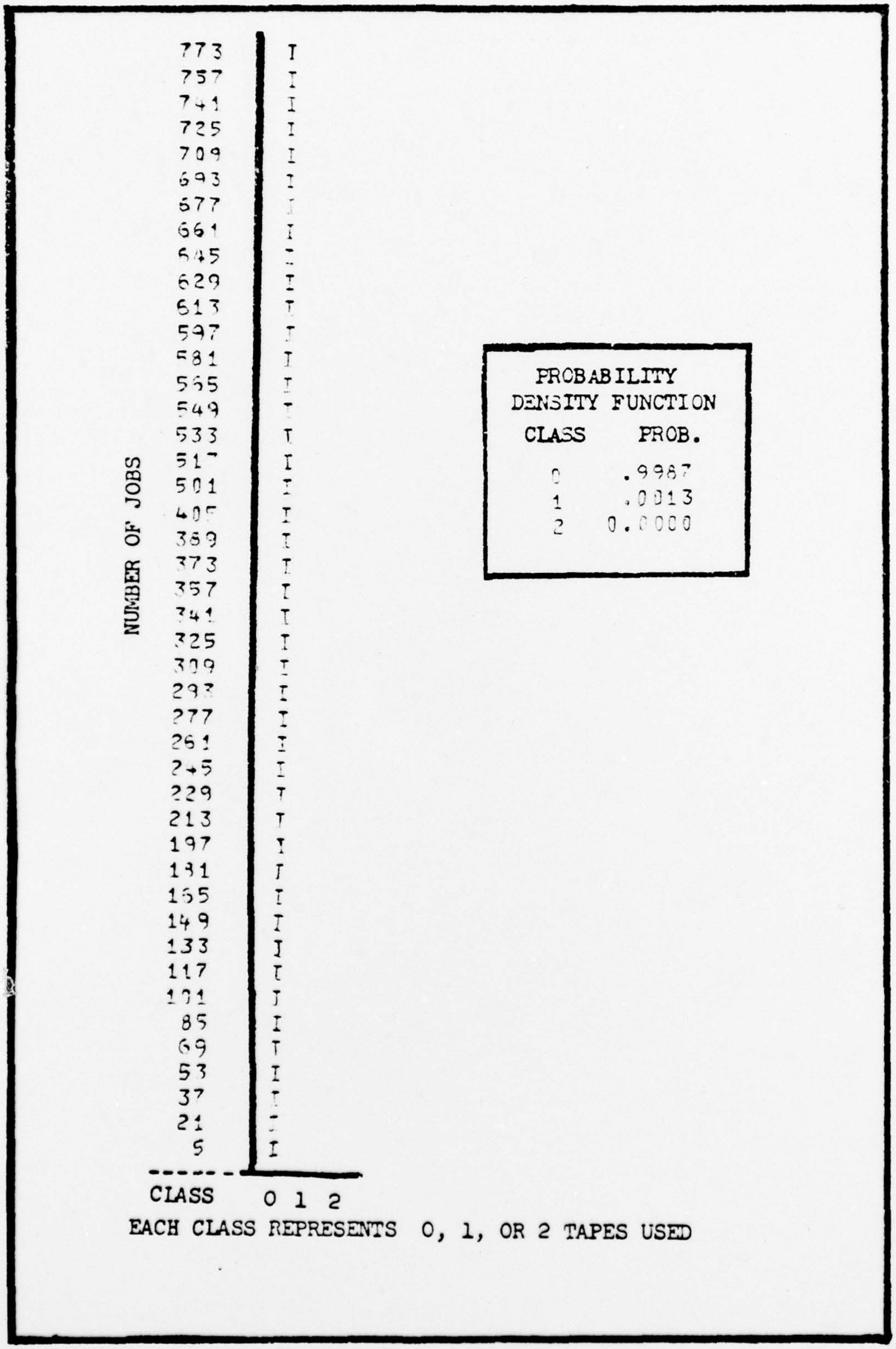


Fig. 37K. Distribution of Tapes Used for 13 October

TABLE XXII

Statistical Summary for Workload Parameters, 13 October

Variable CPTIME CPU time in seconds

Mean	34.095	Std Err	3.573	Std Dev	101.055
Variance	10212.144	Kurtosis	54.495	Skewness	6.723
Minimum	.002	Maximum	1009.873	Sum	27276.382
C.V. Pct	296.389	.95 C.I.	27.082	to	41.109

Variable PPTIME Peripheral Processor time in seconds

Mean	73.833	Std Err	25.840	Std Dev	730.863
Variance	534160.976	Kurtosis	424.569	Skewness	20.205
Minimum	.451	Maximum	16855.554	Sum	59066.332
C.V. Pct	989.888	.95 C.I.	23.111	to	124.555

Variable TIMEIO Input-Output time in seconds

Mean	17.132	Std Err	1.469	Std Dev	41.537
Variance	1725.345	Kurtosis	46.224	Skewness	5.804
Minimum	.074	Maximum	530.549	Sum	13705.536
C.V. Pct	242.456	.95 C.I.	14.249	to	20.015

Variable TOTCOST Total Cost in CRUs

Mean	35.810	Std Err	3.205	Std Dev	90.660
Variance	8219.271	Kurtosis	37.796	Skewness	5.600
Minimum	.040	Maximum	851.326	Sum	28647.621
C.V. Pct	253.173	.95 C.I.	29.518	to	42.101

TABLE XXII (continued)

Variable KWS Memory in kilo-word seconds

Mean	1187.298	Std Err	136.993	Std Dev	3874.749
Variance	1501E+08	Kurtosis	56.417	Skewness	6.892
Minimum	0.000	Maximum	46435.601	Sum	949838.394
C.V. Pct	326.350	.95 C.I.	918.389	to	1456.207

Variable CPOT Control Point Occupancy time in seconds

Mean	1176.287	Std Err	103.427	Std Dev	2925.364
Variance	6557753.717	Kurtosis	31.917	Skewness	5.121
Minimum	1.000	Maximum	29644.000	Sum	941030.000
C.V. Pct	248.695	.95 C.I.	973.266	to	1379.309

Variable CMLOC Central Memory Locations in Use

Mean	129082.880	Std Err	3205.880	Std Dev	90675.974
Variance	.8222E+10	Kurtosis	-1.461	Skewness	.185
Minimum	1024.000	Maximum	261632.000	Sum	.1032E+09
C.V. Pct	70.245	.95 C.I.	122789.938	to	135375.822

Variable CPLOC Control Points in Use

Mean	91.957	Std Err	.357	Std Dev	10.375
Variance	107.648	Kurtosis	25.553	Skewness	-3.675
Minimum	7.000	Maximum	100.000	Sum	73566.000
C.V. Pct	11.283	.95 C.I.	91.237	to	92.678

Variable ROLLOUT Total Time Rolled Out in seconds

Mean	41.821	Std Err	15.929	Std Dev	450.537
Variance	202983.476	Kurtosis	338.764	Skewness	17.772
Minimum	0.000	Maximum	9360.000	Sum	33457.000
C.V. Pct	1077.292	.95 C.I.	10.554	to	73.089

TABLE XXII (continued)

Variable IATIME Inter-arrival time in seconds

Mean	41.491	Std Err	6.427	Std Dev	145.293
Variance	21110.129	Kurtosis	46.948	Skewness	6.499
Minimum	0.000	Maximum	1386.000	Sum	21202.000
C.V. Pct	350.179	.95 C.I.	28.864	to	54.119

Variable INQTIME Time in Input Queue in seconds

Mean	345.601	Std Err	38.643	Std Dev	1054.052
Variance	1111925.958	Kurtosis	39.373	Skewness	5.350
Minimum	0.000	Maximum	12245.000	Sum	257137.000
C.V. Pct	304.991	.95 C.I.	269.738	to	422.464

Variable TAPEREQ Number of Tapes Requested

Mean	.074	Std Err	.010	Std Dev	.276
Variance	.076	Kurtosis	14.472	Skewness	3.788
Minimum	0.000	Maximum	2.000	Sum	59.000
C.V. Pct	373.573	.95 C.I.	.055	to	.093

Variable TAPEUSED Number of Tapes Used

Mean	.071	Std Err	.009	Std Dev	.262
Variance	.069	Kurtosis	11.520	Skewness	3.540
Minimum	0.000	Maximum	2.000	Sum	57.000
C.V. Pct	368.028	.95 C.I.	.053	to	.089

Vita

Jonathan Riggs Bear was born in Kansas City, Missouri, on August 16, 1947. He graduated from the United States Air Force Academy with a Bachelor of Science degree in Computer Science in June of 1969. He began his active duty in the United States Air Force by attending undergraduate pilot training at Craig Air Force Base, Alabama. Upon receiving his pilot wings in 1970, Captain Bear served as a pilot in the EC-121 aircraft at McClellan AFB, California with the 964th Airborn Early Warning and Control (AEW and C) Squadron. From 1972 to 1973, he was assigned with the 23rd Tactical Air Support Squadron in Southeast Asia and flew the OV-10 as a Forward Air Controller and Instructor Pilot. Captain Bear was then assigned to the 963rd AEW and C Squadron and served as an Instructor Pilot and Aircraft Commander of the EC-121. He entered the Air Force Institute of Technology at Wright-Patterson AF Base, Ohio in 1975 to begin study toward a Master of Science degree in Computer Systems. Captain Bear is a member of Tau Beta Pi.

Permanent address: 1006 Terrace Acres Drive
Auburn, Alabama 36830

This thesis was typed by Ms. Carma Troxell.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: GCS/MA/77M-1
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): WORKLOAD CHARACTERIZATION AND PERFORMANCE MEASUREMENT OF THE CDC CYBER 74 COMPUTER SYSTEM
5. TYPE OF REPORT & PERIOD COVERED: MS THESIS
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): JONATHAN R. BEAR
8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Air Force Institute of Technology (AFIT/EN) Wright-Patterson AFB, Ohio 45433
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Mr. James Hudson, Chief of Program Control Br. ASD Computer Center, WPAFB OH 45433
12. REPORT DATE: March 1977
13. NUMBER OF PAGES: 222
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE IAW AFR 190-17.

JERRAL F. GUESS, Capt, USAF
Director of Information

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Workload Characterization
Workload
Computer Performance
Work Measurement
Computer Performance & Evaluation
Job Stream

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In evaluating computer performance, the choice of a specific evaluation approach depends upon the objectives of the investigation, although basically, the methodology involves the observaton and measurement of a computer system or its model while a set of jobs, or workload, is processed. Computer performance evaluation then, can be divided into two focal points, the computer system and its workload. The focus of this thesis is upon the workload.

A methology was developed to process accounting data (continued)

generated by the Control Data Corporation CYBER 74 computer system. A number of statistical tools were used to measure several representative days of accounting data. The workload was characterized by use of a variety of statistical techniques, and the effect of variations in the workload upon computer system performance was measured. The methodology developed for this investigation and the results obtained could be used as a preliminary step in a performance evaluation of the CYBER 74 computer system.